Behavioral/Cognitive

Functional Subdomains within Human FFA

Tolga Çukur,[1,4] **Alexander G. Huth,**[1] **Shinji Nishimoto,**[1] **and Jack L. Gallant**[1,2,3]

[1]Helen Wills Neuroscience Institute, [2]Program in Bioengineering, and [3]Department of Psychology, University of California, Berkeley, California 94720, and
[4]Department of Electrical and Electronics Engineering, Bilkent University, TR-06800 Ankara, Turkey

The fusiform face area (FFA) is a well-studied human brain region that shows strong activation for faces. In functional MRI studies, FFA is often assumed to be a homogeneous collection of voxels with similar visual tuning. To test this assumption, we used natural movies and a quantitative voxelwise modeling and decoding framework to estimate category tuning profiles for individual voxels within FFA. We find that the responses in most FFA voxels are strongly enhanced by faces, as reported in previous studies. However, we also find that responses of individual voxels are selectively enhanced or suppressed by a wide variety of other categories and that these broader tuning profiles differ across FFA voxels. Cluster analysis of category tuning profiles across voxels reveals three spatially segregated functional subdomains within FFA. These subdomains differ primarily in their responses for nonface categories, such as animals, vehicles, and communication verbs. Furthermore, this segregation does not depend on the statistical threshold used to define FFA from responses to functional localizers. These results suggest that voxels within FFA represent more diverse information about object and action categories than generally assumed.

Introduction

One of the most thoroughly studied areas in the human visual cortex is fusiform face area (FFA), a region in the mid-fusiform gyrus that is selective for faces (Sergent et al., 1992; Kanwisher and McDermott, 1997). This area has been reported to produce larger BOLD responses for faces than for objects or houses (Puce et al., 1996; Kanwisher and McDermott, 1997). It has also been argued that FFA responses to faces are relatively more resistant to diverted attention, compared with nonfaces (Reddy and Kanwisher, 2007). These results have led to the hypothesis that FFA is a homogeneous module dedicated to processing a single perceptual category (Kanwisher, 2010).

Nevertheless, the current evidence in support of the modularity hypothesis is not strong enough to draw firm conclusions (Kanwisher, 2010). Several recent studies have raised the possibility that FFA is differentially tuned for object categories other than faces (Gauthier et al., 1999; Ishai et al., 1999; Gauthier, 2000; Haxby et al., 2001; Grill-Spector et al., 2006; Reddy and Kanwisher, 2007; Hanson and Schmidt, 2011; Mur et al., 2012). In particular, it has been reported that multivoxel pattern analyses in FFA can discriminate between nonface object categories, such as shoes versus cars (O'Toole et al., 2005;

Reddy and Kanwisher, 2007). A recent study from our laboratory provides more direct evidence suggesting that at least some FFA voxels might be broadly tuned for both faces and nonface objects (Huth et al., 2012).

The questions regarding category representation in FFA still remain open because no study to date has systematically evaluated voxelwise tuning across a large number of categories to characterize FFA's functional role precisely. Most previous studies of FFA have measured response differences among a small number of object categories. However, this approach does not provide information about categories outside the tested subspace (Friston et al., 2006; Huth et al., 2012), so it is possible that FFA is tuned for other categories as yet untested (Downing et al., 2006). Furthermore, most previous studies have used either ROI analyses or multivoxel pattern analyses. Because these methods aggregate voxel responses across the entire ROI, they are suboptimal for measuring response differences across individual voxels. Thus, it is also possible that FFA contains distinct subregions differing in their category tuning.

Here, we ask whether FFA voxels represent a homogeneous population exclusively selective for a single category (faces), or rather these voxels represent a heterogeneous population with diverse tuning properties. To address this issue, we used fMRI to record BOLD responses elicited by a broad sample of complex natural movies. We then used the WordNet lexical database (Miller, 1995) to label 1705 distinct object and action categories that appeared in the movies (Huth et al., 2012). Next, we used a voxelwise modeling and decoding framework to fit category encoding models to each FFA voxel individually (Naselaris et al., 2009; Huth et al., 2012). Finally, we identified subdivisions within FFA by clustering the voxelwise model weights, and we examined the spatial distribution of the candidate clusters by projecting them onto the cortex.

Materials and Methods

Subjects

The participants for this study were five healthy human subjects, S1–S5 (ages 24–31 years; 4 male, 1 female). All subjects had normal or corrected-to-normal visual acuity. All subjects gave written informed consent before taking part in four separate scan sessions: three sessions for the main experiment and one for the functional localizers. The experimental protocols were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley.

MRI parameters

fMRI data were acquired with a 32-channel receive-only head coil on a 3 T Siemens Tim Trio system (Siemens Medical Solutions) at the University of California, Berkeley. Functional scans were performed using a T_2^*-weighted gradient-echo echo-planar imaging sequence with the following parameters: TR = 2 s, TE = 31 ms, flip angle = 70°, voxel size = $2.24 \times 2.24 \times 3.5$ mm^3, field of view = 224×224 mm^2, and 32 axial slices for whole-brain coverage. The artifacts from fat signal were minimized using a customized water-excitation radiofrequency pulse. Anatomical scans were performed using a T_1-weighted magnetization-prepared rapid-acquisition gradient-echo sequence with the following parameters: TR = 2.30 s, TE = 3.45 ms, flip angle = 10°, voxel size = $1 \times 1 \times 1$ mm^2, field of view = $256 \times 256 \times 192$ mm^3.

Functional localizers

Functional localizer data were collected independently from the main experiment. Following the procedures described by Spiridon et al. (2006), the localizer was acquired within 6 runs of 4.5 min each. A single run consisted of 16 blocks, each lasting 16 s and containing 20 static images drawn from one of the following categories: faces, human body parts, nonhuman animals, objects, spatially scrambled objects, or outdoor scenes. The blocks for separate categories were ordered differently within each run. Individual images were briefly displayed for 300 ms and separated by 500 ms blank periods. To maintain alertness, subjects were instructed to perform a one-back task (i.e., respond with a button press when identical images were displayed consecutively).

Definition of FFA and other ROIs

Standard procedures (Spiridon et al., 2006) were used to functionally define FFA for each individual subject. The localizer runs were first motion-corrected and then registered to the runs from the main experiment. Localizer data were then smoothed in the volume space with a Gaussian kernel of full-width at half-maximum of 4 mm. FFA was denoted as a contiguous cluster of voxels in the mid-fusiform gyrus that responded more strongly to faces than to objects. Recent studies suggest that face and body selective regions neighbor each other in ventral–temporal cortex (Huth et al., 2012; Weiner and Grill-Spector, 2012). To ensure that these body-selective regions did not confound FFA definitions, voxels that responded more strongly to bodies than to faces were removed.

A Student's t test was used to assess the significance of the faces versus objects contrast individually for each voxel. Unfortunately, there is no clear consensus in the literature on the optimal statistical threshold (i.e., p value) for defining functional ROIs, such as FFA (Weiner and Grill-Spector, 2012). Therefore, to ensure that our results were not biased by a particular choice of threshold, we obtained multiple definitions of FFA by selecting different p value thresholds ranging from 10^{-10} to 10^{-4} (uncorrected for multiple comparisons). The overall size of the ROI is enlarged when using relatively less stringent thresholds, but the more liberal definitions always included the voxels assigned to the ROI when using more stringent thresholds (i.e., lower p values). Table 1 lists the numbers of FFA voxels identified using p values between 10^{-6} and 10^{-4} for all subjects individually.

We also defined several other functional ROIs based on standard contrasts (t test, $p < 10^{-5}$, uncorrected). The extrastriate body area (EBA) was defined as the group of voxels in lateral occipital cortex that yielded a positive body versus object contrast. The lateral occipital complex (LOC) was defined as the group of voxels in lateral occipital cortex that yielded a positive object versus spatially scrambled object contrast. The

Table 1. The distribution of FFA voxels across clusters[a]

	Total	Cluster 1	Cluster 2	Cluster3
$p < 10^{-6}$	220	140	80	0
S1	70	43	27	0
S2	67	46	21	0
S3	16	12	4	0
S4	48	24	24	0
S5	19	15	4	0
$p < 10^{-5}$	266	172	94	0
S1	82	51	31	0
S2	82	59	23	0
S3	21	14	7	0
S4	60	31	29	0
S5	21	17	4	0
$p < 10^{-4}$	328	121	129	78
S1	97	32	37	31
S2	103	51	49	9
S3	28	12	8	8
S4	63	15	27	25
S5	23	11	8	5

[a]The numbers of voxels in each cluster are listed for three different localizer thresholds used to define FFA: $p < 10^{-6}$, $p < 10^{-5}$, and $p < 10^{-4}$. The first row in each panel indicates the total number of voxels summed across subjects. The remaining rows indicate the number of voxels in individual subjects (S1–S5).

parahippocampal place area (PPA) was defined as the group of voxels in parahippocampal gyrus that yielded a positive scene versus object contrast. Area MT$^+$ was defined as the group of voxels lateral to the parietal–occipital sulcus that yielded a positive continuous versus scrambled movie contrast. Finally, borders of retinotopic areas were defined using standard techniques (Engel et al., 1997; Hansen et al., 2007).

Natural movie experiment

In the main experiment, BOLD responses were recorded while subjects passively viewed color natural movies presented without sound. To minimize potential stimulus biases, the movies were selected from a wide variety of sources, including the Apple QuickTime HD gallery and www.youtube.com as described by Nishimoto et al. (2011). The original high-definition frames were cropped to a square format and down-sampled to 512×512 pixels ($24° \times 24°$). Subjects fixated on a central spot ($0.16° \times 0.16°$) overlaid onto the movies. To maximize visibility, the color of the fixation point changed randomly at 3 Hz. The stimuli were presented using an MRI-compatible projector (Avotec) and a custom-built mirror system.

Two separate types of datasets were collected: training sets used to fit the voxelwise models, and test sets used to evaluate model predictions. The movies used in the training and test sets did not overlap. The training and test runs were interleaved during each scan session. Training data were collected in 12 runs of 10 min each (a total of 7200 s). Many 10–20 s movie clips were concatenated to construct each run, but each clip appeared only once. Test data were collected in 9 runs of 10 min each (5400 s). The test runs were constructed by first forming 1 min blocks of 10–20 s movie clips and then appending 10 separate blocks. The test blocks were presented in a randomly shuffled order during each run. Each 1 min block was presented a total of 9 times, and the recorded BOLD responses were averaged across repeats. To prevent contamination from hemodynamic transients during movie onset, the last 10 s of the runs were appended to the beginning of the run, and data collected during this period were discarded. The total number of data samples were 3600 and 270 for the training and test runs, respectively. These datasets were previously analyzed (Huth et al., 2012).

Data preprocessing

The FMRIB Linear Image Registration Tool (Jenkinson et al., 2002) was used to align functional brain volumes from individual subjects to correct for changes in head position within and between scan sessions. For each run, a template volume was generated by taking the time average of the motion-corrected volumes within that run. The template volumes for all functional runs (including the localizers) were registered to the first run in the first session of the main experiment. To minimize errors, the
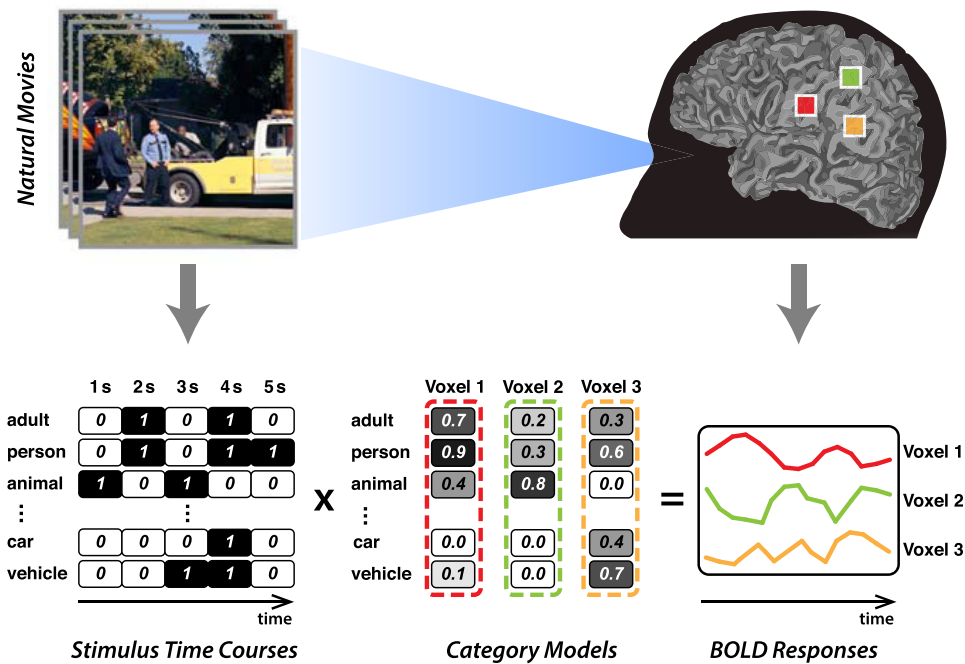
Figure 1. Voxelwise category encoding model. To construct the basis for the category model, the salient object and action categories in each 1 second epoch of movies were labeled using 1705 unique terms from the WordNet lexicon (Miller, 1995). The stimulus time courses were constructed in matrix form, with rows and columns representing distinct categories and epochs, respectively. Regularized linear regression was used to describe individual voxel responses as a weighted sum of these time courses. The fit model weights characterize the category responses of individual voxels to the corresponding object and action categories.

transformations for motion correction and registration were then concatenated, and the original functional data were transformed in a single step.

A median filter with a 120 s window size was used to remove the low-frequency drifts in BOLD responses of individual voxels. The resulting time courses were then normalized to 0.0 ± 1.0 (mean ± SD). The ROI definitions from localizer runs were used to assign voxels to FFA. No spatial or temporal smoothing was applied to the functional data from the main experiment.

Category tuning profiles

We estimated category tuning profiles for each individual voxel in FFA by means of voxelwise modeling and decoding (Gallant et al., 2012). These models effectively describe the relationship between the category content of the natural movies and the evoked BOLD responses (Huth et al., 2012). Therefore, the model weights for each voxel represent its selectivity for a wide variety of categories that appeared in the movies. The following sections describe the selection of the model basis and the regression procedures used to fit these models.

Category encoding model. Before modeling, the WordNet lexicon (Miller, 1995) was used to manually label salient object and action categories in each 1 s period of the movies. WordNet provides a method for disambiguating word senses (e.g., "living organisms lacking locomotion" are called plant.n.02, whereas "industrial buildings" are called plant.n.01), and for capturing hierarchical relationships between these senses. The semantic taxonomy inherent in WordNet was used to infer the presence of more general categories. (For instance, a scene labeled with "cat" must contain a "feline," an "animal," and so on.) The movies were presented without sound, but they contained scenes that depicted verbal communication. To code the presence of communication and to distinguish it from mere presence of people, communication verbs were labeled as action categories. A total of 1705 distinct categories were labeled in the training and test stimuli. For further details on the labeling procedure, see Huth et al. (2012).

Each label can equivalently be represented with a binary indicator variable. These variables were concatenated to form a stimulus time course (categories × seconds) as shown in Figure 1. The voxelwise

models that yielded the best response predictions were determined using linearized regression on the training stimuli and responses. To match the temporal sampling rates of the stimuli and the BOLD responses, the stimulus time course was down-sampled by a factor of 2. To account for the slow hemodynamic BOLD responses, separate linear finite-impulse-response filters were fit to each category. The delays of the finite-impulse-response filters were restricted to 4, 6, and 8 s, or equivalently 2, 3, and 4 samples. The model weights and the finite-impulse-response coefficients were fit simultaneously within the regression procedure.

The voxelwise models were fit using penalized linear regression to prevent overfitting to noise (Fig. 1). The regularization parameter (λ) was selected using a 10-fold cross validation procedure. At each fold, a random 10% of the training data was held out, and the models were fit to the remaining data. The prediction scores were measured by the correlation coefficient (Pearson's r) between the actual and predicted BOLD responses on the held-out set. The optimal λ was determined by maximizing the prediction score of each voxel averaged across all cross-validation folds. Afterward, this λ value was used to refit the models to the entire training set.

The penalty term used during model fitting implicitly reflects expectations about the distribution of responses across 1705 categories in the model. Rather than enforcing strong *a priori* assumptions, we fit two separate models to each voxel using penalty terms based on either the l_2- or the l_1-norm of the model weights. We find that models fit with an l_2-penalty ($R^2 = 0.359 \pm 0.070$, mean ± SD across subjects) explain a greater portion of the response variance than models fit with an l_1-penalty ($R^2 = 0.244 \pm 0.053$, mean ± SD). Therefore, we chose the l_2-penalty to ensure that the fit models would most accurately describe the underlying category representation.

To assess model significance, we evaluated the null hypothesis that the values of the model weights, which characterize the relationship between the stimulus and BOLD responses, can be expected due to chance alone. MRI response measurements are corrupted by an additive white noise that is approximately normally distributed. Thus, we used a Monte Carlo procedure to fit 1000 null models by generating random responses from a standard normal distribution and by using the original stimulus time

course. During this procedure, the zero- and first-order temporal statistics of the noise were matched with those of the BOLD responses for each individual voxel. Furthermore, the original stimulus time course was kept intact to closely match the temporal structures of the null and actual models. The Monte Carlo procedure produced 1000 samples of null model weights for each voxel. Statistical significance of positive weights in the category model was taken as the proportion of samples for which the null model weight was larger than the category model weight. Statistical significance for negative weights was taken as the proportion of samples for which the null model weight was smaller than the category model weight. The significance levels were corrected for multiple comparisons using false discovery rate (FDR) control (Benjamini and Yekutieli, 2001).

To assess model performance, prediction scores of the fit models were computed on the independent test data. Robust estimates of prediction scores were obtained using a jackknifing procedure. The predicted voxelwise BOLD responses on the test data were randomly resampled 10,000 times without replacement (at a rate of 80%). Model performance for each voxel was quantified by averaging prediction scores across jackknife samples. Significance level for each voxel was quantified as the proportion of jackknife samples for which the prediction score was >0 and corrected for multiple comparisons using FDR control. Model fitting procedures were executed using custom software written in MATLAB (MathWorks).

Control models. To assess the significance of FFA responses for nonface categories, a voxelwise face-presence model was fit to the BOLD responses using the procedures described above. The face-presence model contained three separate regressors to account for graded degrees of face presence in the movies. These regressors coded the presence of primary faces (a face is the main object in the movie, and it is highly salient), secondary faces (a face is not the main object, but viewers are likely aware of its presence), and tertiary faces (a face is not the main object, and viewers are likely unaware of its presence). Primary faces were present in 600 s, secondary faces were present in 3560 s, and tertiary faces were present in 465 s of the movies. The full category model was compared with the face-presence model in terms of the proportion of response variance explained in the test data. Significance of differences in explained variance was assessed using bootstrap tests.
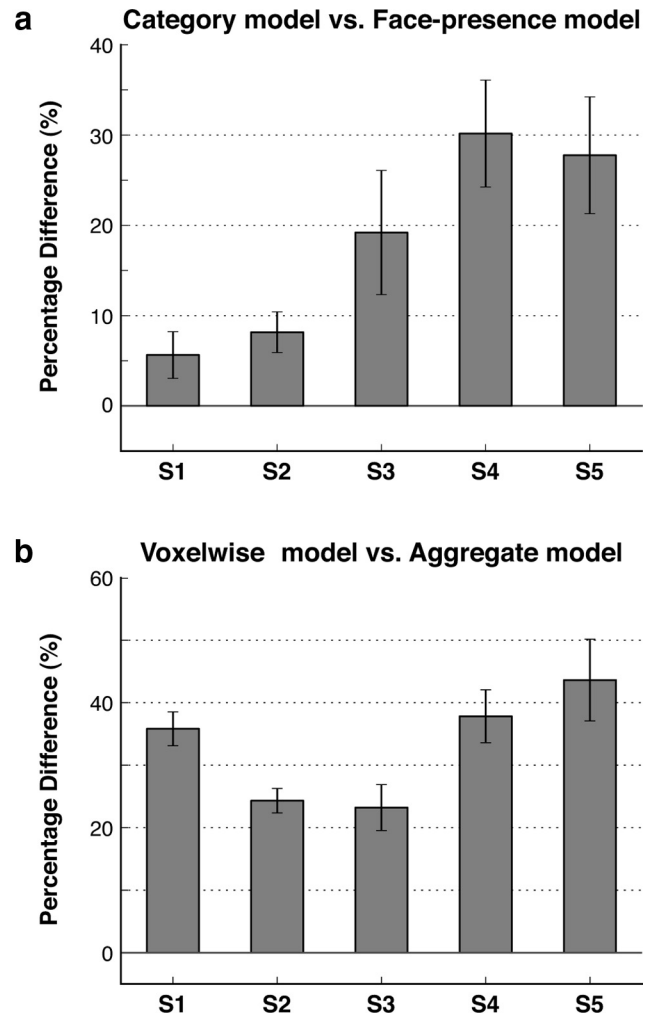
To assess the significance of heterogeneous category tuning in FFA, a separate control analysis was performed using the full category model. First, the prediction score for each voxel was calculated separately using each voxel's own model. Then, the models for the remaining voxels were used to generate response predictions for the held-out voxel, and these predictions were averaged. Each voxel's own model was compared with the average of remaining models in terms of the proportion of response variance explained in the test data. Significance of differences in explained variance was assessed using bootstrap tests.

A separate motion-energy model was used to assess selectivity for simple visual features. It was shown in an earlier study that the motion-energy model accurately predicts voxel responses to natural movies in several retinotopically organized visual areas (Nishimoto et al., 2011). The motion-energy model consisted of 6555 spatiotemporal Gabor wavelet filters. Each filter was constructed by multiplying a 3D spatiotemporal sinusoid by a spatiotemporal Gaussian envelope. Filters occurred at five spatial frequencies (0, 2, 4, 8, 16, and 32 cycles/image), three temporal frequencies (0, 2, and 4 Hz), and eight directions (0, 45 . . . 315 degrees). Filters were positioned on a square grid that covered the movie screen. Grid spacing was determined separately for filters at each spatial frequency so that adjacent Gabor wavelets were separated by 3.5 SDs of the spatial Gaussian envelope.

Clustering analysis

Preliminary inspection of the category tuning profiles revealed a relatively heterogeneous distribution of category selectivity among FFA voxels. Hence, a data-driven clustering approach was adopted to unveil intrinsic group structure among these voxels. All analyses were repeated for multiple FFA definitions based on varying p value thresholds.

Spectral clustering. A spectral clustering algorithm was used that usually offers more reliable performance than traditional techniques (Ng et


a Category model vs. Face-presence model

b Voxelwise model vs. Aggregate model

Figure 2. Significance of tuning for nonface categories and of heterogeneous category tuning. *a*, Separate voxelwise face-presence models were fit to assess the significance of tuning for nonface categories. The face-presence model captured responses to the graded presence of faces in the movies (see Materials and Methods). The full category model (i.e., including 1705 categories) was compared with the face-presence model in terms of response predictions. Bar plots show the percentage difference in explained variance across the population of FFA voxels for subjects S1–S5 (mean ± SEM). The full model increases the explained variance in each subject ($p < 10^{-4}$, bootstrap test). This result indicates that tuning for nonface categories is significant. *b*, A separate control analysis was performed to assess the significance of heterogeneous category tuning within FFA. The full category model for each voxel was compared with an aggregate model obtained by averaging the voxelwise models across the remaining voxels. Bar plots show the percentage difference in explained variance across the population of FFA voxels for subjects S1–S5 (mean ± SEM). The voxelwise models increase the explained variance in each subject ($p < 10^{-4}$, bootstrap test). This result indicates that the heterogeneity of category tuning among FFA voxels is significant.

al., 2001; Zelnik-Manor and Perona, 2004). The spectral clustering algorithm first forms a fully connected affinity graph $G = (V, E)$ with vertices corresponding to individual tuning profiles. The edges connecting these vertices are assigned non-negative weights, W, that represent the affinity between the curves. Here, we used a reliable self-tuning affinity measure as follows:

$$W_{ij} = \begin{cases} e^{-d_{ij}^2 / 2\sigma_i \sigma_j}, & i \neq j \\ 0, & i = j \end{cases} \tag{1}$$

where σ_i denotes the scale parameter equal to the average distance between the curve for the ith voxel and the curves for all other voxels (Zelnik-Manor and Perona, 2004).

Similarity CDFs   **Differences**
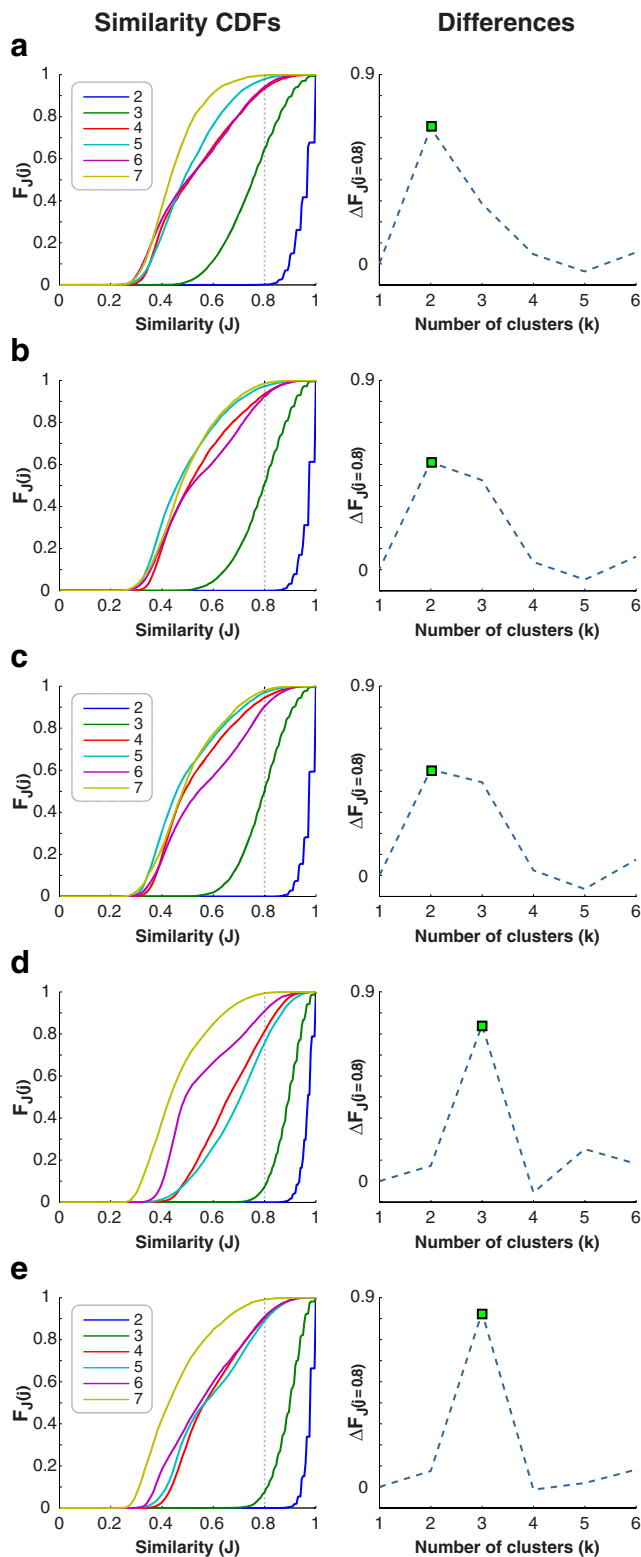
Figure 3. The number of clusters. A stability-based validation technique was used to determine the optimal number of clusters for multiple FFA definitions based on varying localizer thresholds: **a**, $p < 1 \times 10^{-6}$. **b**, $p < 5 \times 10^{-6}$. **c**, $p < 1 \times 10^{-5}$. **d**, $p < 5 \times 10^{-5}$. **e**, $p < 1 \times 10^{-4}$. For each FFA definition, the cluster stability was determined separately for several numbers of clusters (k). This validation technique measures the cluster stability against random perturbations in the dataset. Specifically, stability was measured as the similarity of clustering solutions that are performed on random subsamples (i.e., 80%) of FFA voxels pooled across subjects. The pairwise similarities (J) between the clustering solutions were histogrammed to generate the cumulative distribution functions (CDFs) shown in the left column. If the solutions are stable, then the distributions will be concentrated around 1. If the solutions are instead

Afterward, spectral clustering attempts to partition the graph such that the resulting clusters are as weakly connected as possible (i.e., the connecting edges have low weights) (Luxburg, 2007). Ideally, each cluster denotes a subset of vertices that are connected to each other by a series of edges and disjoint from the remaining vertices in the graph. Such clusters can be found by computing the eigenvalue decomposition of the graph Laplacian matrix, $L \in R^{n \times n}$ where n is the number of voxels (Ng et al., 2001). This matrix has as many zero-valued eigenvalues as there are clusters in the partitioned graph. The corresponding eigenvectors indicate the vertices grouped in each connected cluster.

For a given number of clusters k, the clustering algorithm retrieves the eigenvectors corresponding to the k smallest eigenvalues, and appends them to form a new matrix $U \in R^{n \times k}$. This dimensionality-reduced representation emphasizes the group structure in the data, such that the next step of the spectral clustering algorithm, a simple k-means clustering applied on the rows of U, can trivially detect the clusters (MacQueen, 1967). Because k-means clustering can be sensitive to the choice of initial cluster centers, the clustering analyses were performed with 20 different random initializations. The initialization that yielded the lowest sum-of-squares of within-cluster distances was selected as the optimal solution. This procedure was repeated 10 times, and it was observed that 20 random initializations are sufficient to obtain the same optimal solutions across repetitions.

Distance function. Similar to many other clustering methods, spectral clustering requires the selection of a distance function to characterize the dissimilarity between the data points. A correlation-based measure was used to quantify the distance between pairs of tuning profiles (DeAngelis et al., 1999):

$$d_{ij} = 1 - \frac{a_i \cdot a_j}{\|a_i\| \|a_j\|} \tag{2}$$

Here, d_{ij} is the distance between tuning profiles of voxels i, j, a_i, and a_j are the corresponding 1705-dimensional vectors denoting the tuning profiles, $\|\cdot\|$ indicates the l_2-norm, and $\cdot$ is a dot-product operation.

Number of clusters. Finding the optimum number of clusters is an important parameter selection problem for many clustering algorithms. A stability-based validation technique was used to quantitatively assess the quality of the clustering results, and to rationally determine the number of clusters without relying on external information (Ben-Hur et al., 2002; Handl et al., 2005). This technique performs clustering analyses on subsamples of the data, and uses pairwise similarities between the clustering solutions as a measure of stability. Different numbers of clusters are then compared through their corresponding distributions of clustering stability.

For different numbers of clusters k (1,2. . . 7), a separate jackknifing procedure was performed with 5000 iterations to estimate the probability distribution of stability. (No computations were required for $k = 1$, as the tuning profiles will always be placed in the same cluster.) At each iteration, 80% of the curves were randomly drawn twice, and spectral clustering was performed separately on these two subsamples. The clustering results for the common set of curves included in both subsamples were characterized with labeling matrices, C as follows:

$$C_{ij} = \begin{cases} 1, & a_i \text{ and } a_j \text{ are in the same cluster} \quad \& \quad i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

←

unstable, the distributions will be widespread. Thus, the optimal number of clusters can be identified by a sudden transition from narrow to more widespread distributions. To detect such transitions, we calculated the percentage of similarity measurements that were below a high similarity threshold (i.e., J < 0.8) for each number of clusters. The differences in this percentage value between consecutive numbers of clusters are plotted in the right column. The optimal number of clusters was determined as the point where the difference in maximized (green dot).
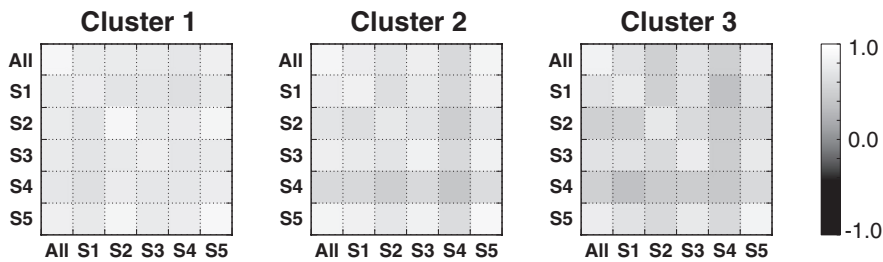
Figure 4. Similarity of cluster centers across subjects. To examine whether the functional subdomains within FFA are consistent across subjects, clustering analysis was performed for each subject individually. The cluster centers computed in individual subjects were compared across subjects and compared with the group cluster centers. Similarity measurements were performed between cluster centers estimated separately from responses to the first and second halves of the movies. FFA voxels were defined using a localizer threshold of $p < 10^{-4}$. The similarities (as measured by correlation) of cluster centers are displayed with separate matrices. The row and column labels identify the group (All) and individual-subject (S1, S2, S3, S4, and S5) cluster centers. The color scale ranges from black for negative correlation (-1) to white for positive correlation (1). The group and individual-subject cluster centers are quite similar ($p < 10^{-4}$, bootstrap test). This indicates that the functional heterogeneity within FFA is reliable for individual subjects and consistent across subjects.

The similarity of the two clustering results, $C^{(1)}$ and $C^{(2)}$, were then computed using the Jaccard Index (Jaccard, 1908):

$$J(C^{(1)}, C^{(2)}) = \frac{C^{(1)} : C^{(2)}}{C^{(1)} : C^{(1)} + C^{(2)} : C^{(2)} - C^{(1)} : C^{(2)}} \quad (4)$$

here ":" denotes the Frobenious innerproduct for matrices. The probability distribution functions $P(j)$ and thereby cumulative distribution functions $F_J(j)$ were obtained by calculating the normalized histograms (with a bin width of 0.005) of 5000 similarity values.

Finally, the optimal number of clusters was determined from the resulting cumulative distribution functions (Ben-Hur et al., 2002). The optimal value can be identified by a sudden transition from distributions concentrated around a similarity of 1 (i.e., stable clustering solutions) to more widespread distributions (i.e., unstable clustering solutions). We used a quantitative rationale to detect such transitions as follows:

Rationale
For each distribution of similarity $F_J(j) = P(J < j)$, determine the value of $F_J(0.8) = P(J < 0.8)$.

Compute the finite differences of these values across consecutive numbers of clusters (k), $\Delta F_J^k = F_J^{k+1}(0.8) - F_J^k(0.8)$.

The optimal k is chosen to maximize the finite difference ΔF_J^k.

Unstable clustering solutions will generate more uniform distributions of similarity, and $P(J < 0.8)$ will be relatively small. In contrast, stable solutions will have narrower distributions and yield higher $P(J < 0.8)$. Therefore, there will be a relatively big jump in $P(J < 0.8)$ values during the transition from stable to unstable solutions. This technique will successfully detect the absence of clusters in the data, as in this case the distributions for $k > 1$ will all be considerably widespread.

Repeatability. By selecting the optimal number of clusters based on stability, a stringent criterion is enforced on the repeatability of the clustering results. This criterion assures that clustering labels are reliable against perturbations in the dataset. However, the stability of labels does not provide direct information about the variability in the cluster centers, which is critical for the interpretation of the results. Such variability may arise from the k-means step in spectral clustering that relies on random initializations. Therefore, the clustering analysis may potentially yield different cluster centers in each run.

To assess the repeatability of the cluster centers, the spectral clustering analysis was repeated 10,000 times using the entire dataset (instead of subsamples) and the optimal number of clusters determined in the previous step. Apart from variations in the centers resulting from k-means, the cluster labels can also be randomly permuted in between iterations. Therefore, the corresponding cluster labels were matched by maximizing the overlap between the identities of the voxels assigned to each cluster. In other words, clusters that

shared the largest number of voxels were matched. The similarity between matched pairs of cluster centers were then quantified by correlation coefficients (Pearson's r). Finally, bootstrap tests were performed to determine whether these correlations were significantly above a stringent threshold of 0.95.

Visualization in model space
To interpret the results of the clustering analysis, the cluster centers were individually visualized with graphs in the model space. These graphs were constructed by plotting the objects and actions in our category model with separate tree structures. Each graph vertex was used to denote a distinct feature in the category model. Meanwhile, graph edges were used to represent the hierarchical WordNet relations between superordinate and subordinate categories. The locations of the vertices and the length of the edges were manually assigned for ease of visualization and do not convey information about semantic distances between categories. The size and color of the vertices indicate the magnitude and sign of the predicted responses to the corresponding categories, respectively.

Visualization on cortical flatmaps
To examine the organization of category representation across the cortex, flatmaps of the cortical surface were generated from anatomical data. The category tuning profiles were then visualized on cortical flatmaps. For this purpose, a principal components analysis was first used to recover the dimensions of the semantic space represented by FFA. The tuning profile of each voxel was then projected into this semantic space. Statistical significance of the projections was evaluated using the null models generated by the aforementioned Monte Carlo procedure (see Category encoding model). The null models for each voxel were also projected into the semantic space. To improve the quality of the semantic space, the analysis was restricted to the first 10 principal components that individually explain >1% of the response variance. The category tuning of a voxel was deemed statistically significant if its projection into the semantic space was significantly separated from the projections of the null models (χ^2 test based on Mahalanobis distance). The significance levels were corrected for multiple comparisons using FDR control.

Spatial distribution of clusters
To examine whether FFA clusters are spatially segregated, spatial locations of FFA voxels were measured separately in the volumetric brain spaces of each subject. If clusters are spatially segregated, pairwise distances between voxels should be smaller for voxels within the same cluster than for voxels in different clusters. The distribution of within- versus across-cluster distances was measured separately for each cluster. This distribution was compared with a null distribution obtained by randomly assigning voxels to three clusters. Significance of spatial segregation was assessed using bootstrap tests.

To assess characteristic differences in the spatial distribution of individual FFA clusters, two functional ROIs that lie along the ventral collateral sulcus (CoS) and middle temporal sulcus (MTS), namely, PPA and EBA, were used as spatial landmarks. Voxelwise spatial locations were measured separately in the volumetric brain spaces of individual subjects. The centers of PPA and EBA were obtained by averaging the spatial locations of entailed voxels. A reference direction that linearly traverses from the center of EBA to PPA was computed in the volumetric space. The spatial location of each FFA voxel was projected onto this reference direction. Voxels that are spatially closer to PPA should have larger projections, and voxels that are spatially closer to EBA should have smaller projections. Significance of differences in spatial locations was assessed using bootstrap tests.
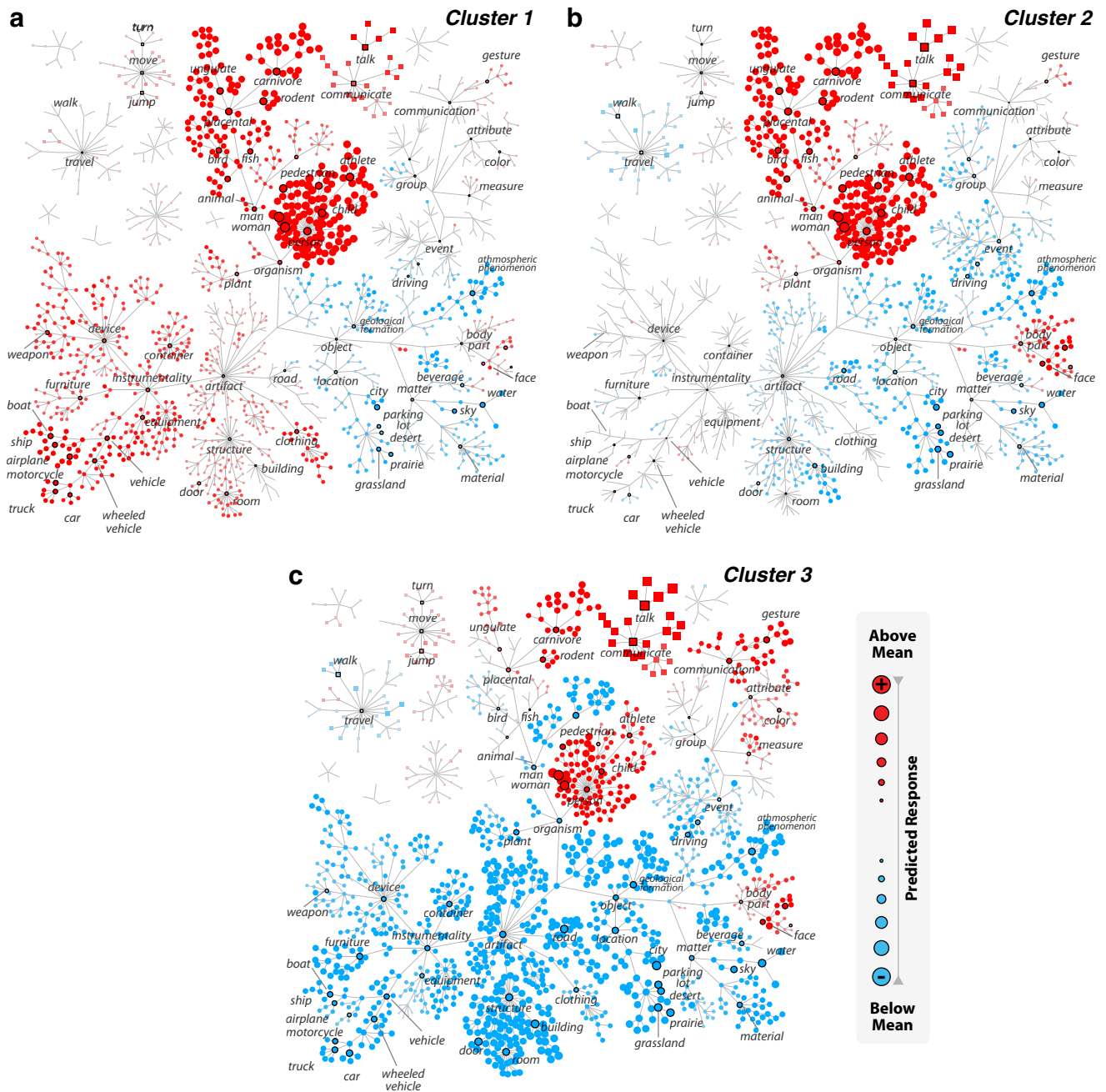
Figure 5. The cluster centers. Spectral clustering analysis among FFA voxels reveals three functional subdomains. ***a– c,*** Mean tuning profiles across voxels within each cluster are shown using a series of graphs for object (main tree, circular vertices) and action (smaller trees, square vertices) categories. Subsets of the categories are labeled to orient the reader. The size of each vertex indicates the magnitude, whereas its color indicates the sign (red represents $+$; blue represents $-$) of the category response (see legend). ***a,*** Responses of the first cluster are strongly enhanced by humans and animals and weakly enhanced by man-made instruments, including vehicles ($p < 0.05$, Monte Carlo test, FDR corrected). ***b,*** Responses of the second cluster are strongly enhanced by humans and animals ($p < 0.05$) and weakly enhanced by body parts and communication verbs, including primary faces (faces that are the most salient object in the scene). ***c,*** Responses of the third cluster are strongly enhanced by humans, placental mammals, communication verbs, gestures (i.e., facial gestures), and faces but strongly suppressed by man-made artifacts, particularly structures, such as buildings and rooms ($p < 0.05$). Responses of all three clusters are suppressed by many natural objects, such as geological landscapes and natural materials ($p < 0.05$).

Silhouette width

The separability of the identified clusters was calculated using the average silhouette width (Peter, 1987). This average width reflects the quality of clustering and is quantified as the mean of voxelwise silhouette widths S_i:

$$S_i = \frac{\min_{Z \neq Z_i}(\bar{d}_{ij})_{j \epsilon Z} - (\bar{d}_{ij})_{j \epsilon Z_i}}{\max\left(\min_{Z \neq Z_i}(\bar{d}_{ij})_{j \epsilon Z}, (\bar{d}_{ij})_{j \epsilon Z_i}\right)} \quad (5)$$

Here Z denotes the set of voxels assigned to a given cluster and Z_i denotes cluster label for the i th voxel. $(\bar{d}_{ij})_{j \epsilon Z}$ is the average distance between the i th voxel and all voxels in cluster Z.

Results

Category tuning of FFA voxels

The human FFA is commonly assumed to be involved in category representation (Kanwisher, 2010). To measure FFA tuning for hundreds of object and action categories, we fit separate category
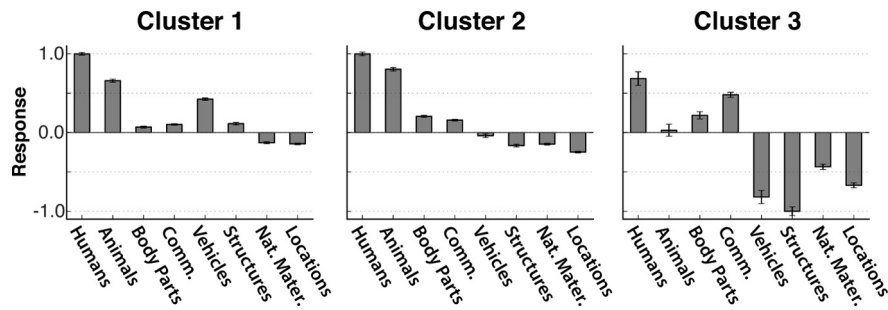
Figure 6. Response levels to object and action categories in each cluster. The responses of each cluster to eight distinct categories were computed, including humans, animals, body parts, communication, vehicles, structures (e.g., building, room), natural materials (e.g., water, soil), and geographic locations (i.e., mountain, city). The response to each category was computed as the mean response to all of its subordinate categories. Individual voxel responses were averaged within each cluster to compute the average category responses (mean ± SEM, across voxels within each cluster). The first cluster is broadly tuned for humans, animals, and vehicles. The second cluster is tuned for humans and animals. Finally, the third cluster is tuned for humans, body parts, and communication.

models to single FFA voxels in each individual subject (Fig. 1). FFA definitions were obtained from functional localizer data by first removing voxels with a positive body-versus-face contrast and then applying varying p value thresholds to a face-versus-object contrast (see Materials and Methods). We find that the category models provide significant response predictions for all FFA voxels ($p < 10^{-4}$, jackknife test, FDR corrected).

Evidence from several fMRI studies implies that FFA might be differentially tuned for nonface objects (O'Toole et al., 2005; Reddy and Kanwisher, 2007; Huth et al., 2012). If FFA exhibits tuning for nonface categories, then nonface category responses should account for a significant portion of the total response variance in FFA voxels. To assess whether FFA is significantly tuned for nonface categories, we performed two complementary analyses. First, we separately fit a face-presence model to each FFA voxel that captured responses to the graded presence of faces in the movies (see Materials and Methods). We reasoned that, if nonface category responses are significant, the full category model should yield better predictions of the BOLD responses than the face-presence model. Figure 2a shows the difference in variance explained by the full category model versus the face-presence model. We find that the proportion of explained variance across FFA voxels is $R^2 = 0.359 \pm 0.070$ (mean ± SD across subjects) for the full category model and $R^2 = 0.308 \pm 0.074$ for the face-presence model. The average improvement in explained variance across the population of FFA voxels is $18.2 \pm 9.6\%$ (mean ± SD across subjects). In every subject, the difference in explained variance is significant ($p < 10^{-4}$, bootstrap test). Thus, tuning for nonface categories accounts for a significant and substantial part of the response variance in FFA.

Next, we performed a separate analysis to ensure that the additional response variance explained by the full category model over the face-presence model is not an artifact of scenes that were correlated with the presence of faces. For this purpose, we refit the full category model after removing all training data (including a 6 s safety margin to account for hemodynamic delays) collected while humans, animals, or body parts were present in the movies regardless of face presence. We reasoned that, if the difference in explained variance between the category and face-presence models is the result of nonface category responses, then the category model fit to face-absent data should explain a comparable proportion of variance to this difference. Indeed, we find that the proportion of variance explained by the category model fit to face-absent data is $R^2 = 0.058 \pm 0.034$ (mean ± SD across sub-

jects; $p < 10^{-4}$, bootstrap test), which is nearly identical to the additional variance explained by the category model fit to all training data ($\Delta R^2 = 0.051 \pm 0.021$, mean ± SD). This finding affirms that FFA tuning for nonface categories is not an artifact of subjective face percepts.

One recent fMRI study proposed that category tuning might be heterogeneously distributed across voxels within FFA (Grill-Spector et al., 2006). If FFA exhibits spatially heterogeneous category tuning, then intervoxel differences in category responses should account for a significant portion of the total variance in FFA voxels. In turn, this would imply that each voxelwise model should yield better predictions of that particular voxel's responses compared with an aggregate model averaged across the remaining voxels. Figure 2b shows the difference in variance explained by each voxel's own model versus the average FFA model. We find that the proportion of explained variance across FFA voxels is $R^2 = 0.359 \pm 0.070$ (mean ± SD across subjects) for each voxel's own model and $R^2 = 0.266 \pm 0.049$ for the average FFA model. In every subject, the voxelwise model explains significantly more variance than the average FFA model ($p < 10^{-4}$, bootstrap test). The average improvement in explained variance across the population of FFA voxels is $34.4 \pm 4.6\%$ (mean ± SD across subjects). This improvement cannot be attributed to simple response baseline or gain differences among FFA voxels because BOLD responses of each voxel were individually z-scored before modeling and because prediction scores were measured using correlation. Thus, this result demonstrates that differences in category tuning among FFA voxels account for a significant portion of FFA responses.

Clustering analysis

To systematically assess variations of category tuning within FFA, we used spectral clustering to analyze voxelwise tuning profiles in individual subjects. A stability-based technique was used to select the optimal number of clusters (k) without relying on external information (see Materials and Methods for details). The optimal k was determined separately for each FFA definition based on varying localizer thresholds. We find that the FFA voxels fall into three distinct clusters for every subject, even at stringent thresholds (starting with $p < 5 \times 10^{-5}$; see Table 1). Even at exceedingly stringent thresholds down to $p < 10^{-10}$, FFA voxels fall into two distinct clusters. Typical thresholds used to define FFA in previous studies range from 10^{-5} to 10^{-3} (Spiridon and Kanwisher, 2002; Grill-Spector et al., 2006; Reddy and Kanwisher, 2007). To increase the sensitivity in verifying the optimal number of clusters, we repeated the stability-based validation at the group level after pooling FFA voxels across subjects. We find that the FFA voxels fall into three distinct clusters at the group level, consistent with the results obtained at the level of individual subjects (Fig. 3). These results indicate that FFA has at least three functional subdomains for category tuning.

To examine the consistency of FFA subdomains across subjects, we compared the category tuning of the voxel clusters identified in separate subjects. The category tuning of each cluster was taken as the mean tuning profile across FFA voxels within the cluster (i.e., the cluster center). The stability of the cluster centers
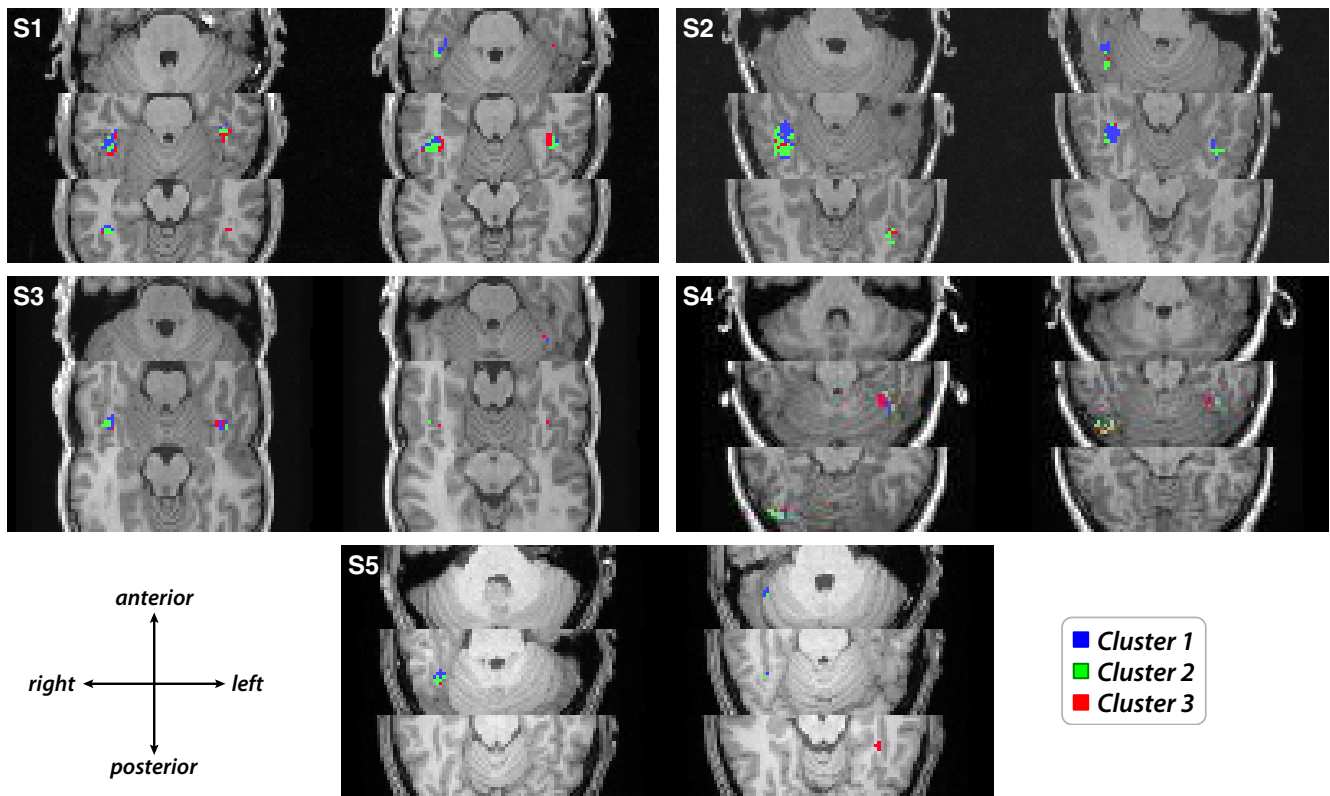
Figure 7. Spatial distribution of clusters. The spatial distribution of clusters shown in the volumetric brain spaces of individual subjects (S1–S5). FFA voxels were defined using a localizer threshold of $p < 10^{-4}$. Separate clusters are labeled with blue, green, and red colors overlaid onto anatomical images (see legend). Consecutive axial slices at the ventral occipitotemporal areas are grouped in montage format. The clusters are spatially segregated in the brain space of each subject ($p < 10^{-5}$ for S1–S3 and S5, $p < 0.02$ for S4, bootstrap test). The first cluster that has the strongest tuning for instruments, vehicles, and structures among the three clusters is located anteriorly relative to the remaining clusters, and closer to regions along the ventral collateral sulcus that are assumed to be selective for scenes with man-made structures ($p < 10^{-4}$, bootstrap test). The second cluster that has the strongest tuning for humans and animals is located posteriorly and closer to regions along the middle temporal sulcus that are known to be selective for animate motion and bodies ($p < 10^{-4}$). Last, the third cluster that has the strongest tuning for communication verbs and human body parts and weakest selectivity for animals is located centromedially ($p < 10^{-4}$). There is no significant hemispheric lateralization for any of the three clusters ($p > 0.34$, bootstrap test).

was first evaluated using a bootstrap procedure (see Materials and Methods). We find that all three cluster centers are highly stable and repeatable in individual subjects ($p < 10^{-4}$, bootstrap test). The cluster centers were then compared across subjects. We find that the cluster centers are highly correlated across subjects ($r = 0.73 \pm 0.15$, mean $\pm$ SD across subjects, $p < 10^{-4}$, bootstrap test). We also find that the within-cluster correlations are higher than the across-cluster correlations ($\Delta r = 0.13 \pm 0.04$, $p < 10^{-4}$).

To facilitate intersubject comparisons and increase sensitivity, we repeated the clustering analyses at the group level. We find that all three group cluster centers are highly repeatable ($p < 10^{-4}$, bootstrap test) and that they are highly correlated with individual-subject cluster centers ($r = 0.86 \pm 0.15$, mean $\pm$ SD across subjects, $p < 10^{-4}$, bootstrap test).

To ensure that intersubject consistency of the clusters is not biased by spurious correlations in the movies, we also measured the similarity of cluster centers estimated separately from responses to the first and second halves of the movies. We find that the split-half cluster centers are highly correlated across subjects (Fig. 4; $r = 0.65 \pm 0.14$, mean $\pm$ SD, $p < 10^{-4}$, bootstrap test) and with the group cluster centers ($r = 0.73 \pm 0.14$, mean $\pm$ SD). Together, these results suggest that functional heterogeneity among FFA voxels is reliable for individual subjects and consistent across subjects.

Finally, to reveal the aspects of category information that are represented in each cluster, we assessed the differences in category tuning across the clusters. For this purpose, we inspected the three group cluster centers obtained when FFA was defined using a localizer threshold of $p < 10^{-4}$ (Fig. 5; see Fig. 13 for individual-subject cluster centers; see Fig. 14 for raw model weights of the group cluster centers). To highlight the key differences between the cluster centers, we compared the average response levels of each cluster with several important object and action categories (Fig. 6). BOLD responses of the first cluster are strongly enhanced by humans and animals and weakly enhanced by man-made instruments, including vehicles ($p < 0.05$, Monte Carlo test, FDR corrected). Responses of the second cluster are strongly enhanced by humans and animals ($p < 0.05$) and weakly enhanced by communication verbs and body parts, including primary faces (faces that are the most salient object in the scene). Responses of the third cluster are strongly enhanced by humans, placental mammals, communication verbs, gestures (i.e., facial gestures), and faces but strongly suppressed by man-made artifacts, particularly structures, such as buildings and rooms ($p < 0.05$). Responses of all three clusters are suppressed by many natural objects, such as geological landscapes and natural materials ($p < 0.05$). These results suggest that the clusters in FFA are broadly tuned for many nonface categories, and differ in their tuning for these nonface categories.
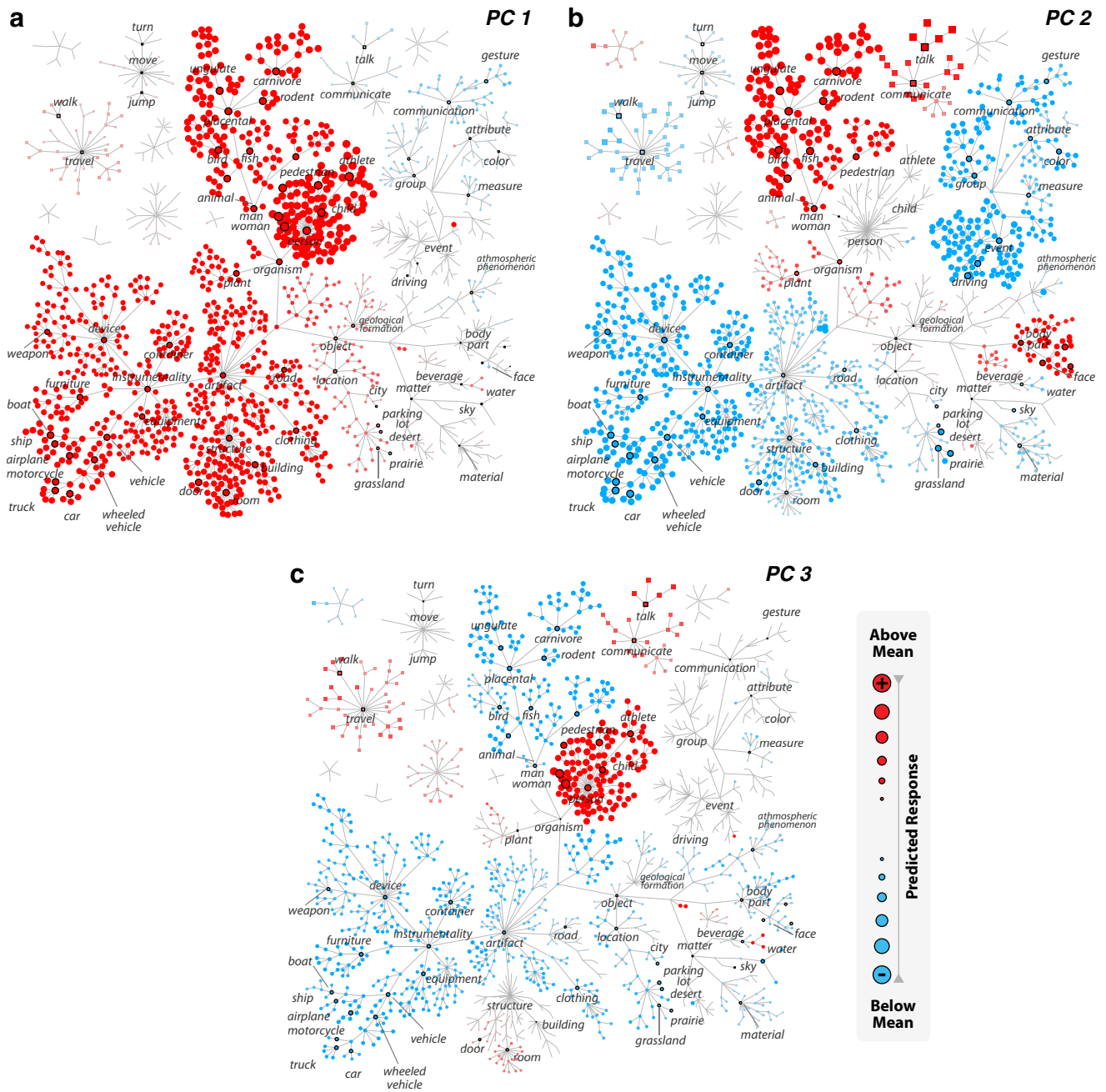
Figure 8. PCA. To assess the important dimensions of category tuning, a PCA was performed across tuning profiles of FFA voxels. The tuning profiles were pooled across subjects before PCA. The first three PCs account for 55.8 ± 8.6% (mean ± SD across subjects) of the variance in category responses of FFA voxels (first PC, 32.1%; second PC, 12.0%; third PC, 11.9%). These PCs are presented in separate graphs, each formatted as in Figure 5. **a**, The first PC is broadly tuned for categories that appear in civilized contexts, such as humans, animals, vehicles, indoor spaces, roads, and devices ($p < 0.05$, Monte Carlo test, FDR corrected). **b**, The second PC distinguishes categories related to living things and their interactions, such as animals, humans, body parts, and communication verbs from nonliving categories, such as vehicles, indoor spaces, devices, and events ($p < 0.05$). **c**, The third PC distinguishes people from animals and nonliving categories, including vehicles and devices ($p < 0.05$).

Functional organization within FFA

It is thought that relatively early stages of the human visual cortex are functionally organized such that neural populations with similar tuning profiles are spatially clustered (Sereno et al., 1995). To assess whether an analogous principle governs the organization of voxels within human FFA, we examined the spatial distribution of voxel clusters in the volumetric brain spaces of individual subjects (Fig. 7). We find that the voxel clusters identified only by category tuning are also spatially segregated in the brain space of each subject ($p < 10^{-5}$ for S1–S3 and S5, $p < 0.02$

for S4, bootstrap test; see Materials and Methods). The first cluster, which has the strongest tuning for instruments, vehicles, and structures among the three clusters, is located anteriorly relative to the remaining clusters and closer to regions along ventral CoS that are assumed to be selective for scenes with man-made structures ($p < 10^{-4}$ group level, $p < 0.35$ subject level, bootstrap test). The second cluster, which has the strongest tuning for humans and animals, is located posteriorly and closer to regions along MTS that are known to be selective for animate motion and bodies ($p < 10^{-4}$ group level, $p < 0.005$

subject level). The third cluster, which has the strongest tuning for communication verbs and human body parts and weakest tuning for animals, is located centromedially ($p < 10^{-4}$ group level, $p < 0.01$ subject level). These results clearly indicate that voxels that exhibit similar category tuning are located in close spatial proximity within FFA.

A recent study from our laboratory suggested that the brain represents visual categories as distinct locations in a continuous semantic space (Huth et al., 2012). To recover the semantic space represented by FFA and to assess whether the dimensions of this space are consistent across subjects, we performed a separate principal component analysis (PCA) on voxelwise category tuning profiles. Our category model spans a 1705-dimensional space of category responses. Thus, PCA on tuning profiles recovers important axes of category tuning in this space that are represented across FFA voxels. To evaluate the consistency of the PCs across subjects, we measured the similarity between the PCs that were separately estimated from responses to the first and second halves of the movies. We find that the first three PCs of FFA voxels estimated within individual subjects are highly correlated across subjects ($r = 0.58 \pm 0.04$, $p < 10^{-4}$, bootstrap test). These individual-subject PCs are also highly correlated with the group PCs of FFA voxels (0.61 ± 0.05, $p < 10^{-4}$), explaining $55.8 \pm 8.6\%$ (mean $\pm$ SD across subjects) of the variance in category responses of FFA voxels. These results imply that category representation in FFA is organized according to at least three semantic dimensions that are highly consistent across subjects.

To reveal which specific aspects of semantic information are reflected in the PC dimensions, we inspected the first three group PCs of FFA voxels (Fig. 8; see Figure 15 for raw model weights of the PCs). The first PC is broadly tuned for categories that appear in civilized contexts, such as humans, animals, vehicles, indoor spaces, roads, and devices ($p < 0.05$, Monte Carlo test, FDR corrected). The second PC distinguishes categories related to living things and their interactions, such as animals, humans, body parts, and communication verbs from nonliving categories, such as vehicles, indoor spaces, devices, and events ($p < 0.05$). The third PC distinguishes people from animals and nonliving categories, including vehicles and devices ($p < 0.05$). These results imply that FFA represents information about the context and animacy of object and action categories.

Our analyses reveal three semantic PCs that are highly consistent across subjects, and these PCs explain a substantial portion of the response variance across FFA voxels. These findings suggest that the FFA clusters designate voxel populations that represent separate regions within the semantic space (Huth et al., 2012). Thus, we hypothesized that the voxel clusters should be distinguishable in the space defined by the semantic PCs. To test this hypothesis, we projected the category tuning profiles of FFA voxels onto the first two PCs (Fig. 9). We find that the three clusters of FFA voxels project to segregated regions in the PC space ($p < 10^{-4}$, bootstrap test). This result independently validates the clustering analyses, and it indicates that the voxel clusters reflect the organization of FFA along multiple dimensions of the semantic space.

Functional organization outside FFA

It remains an open question whether the functional organization of FFA is similar to the organization of other category-selective regions in occipitotemporal cortex (Kanwisher, 2010). Several recent studies have implied that FFA might be embedded in a larger scheme of category representation across high-level visual cortex (Haxby et al., 2011; Huth et al., 2012). We therefore asked
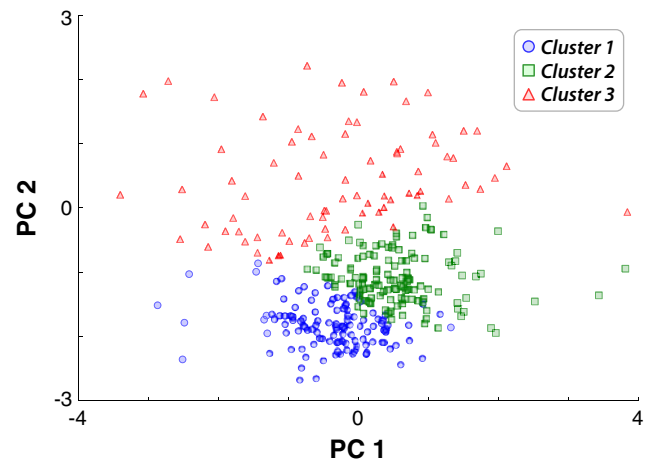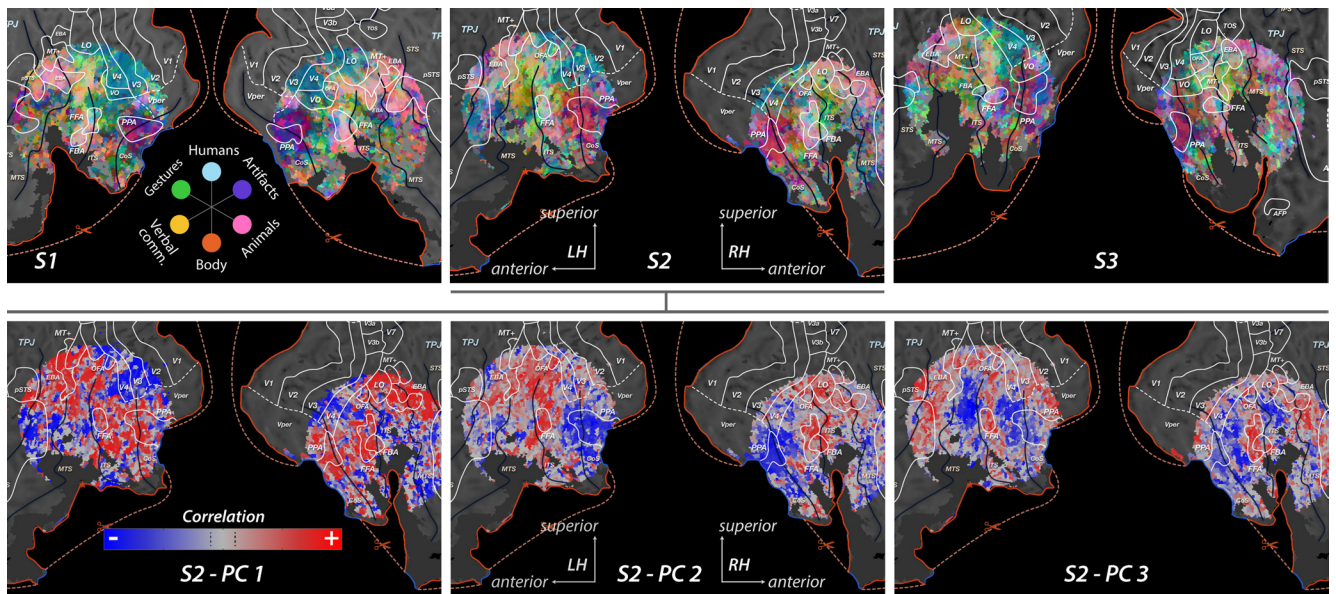

Figure 9. Projections of voxelwise tuning profiles onto PCs. To independently assess the functional heterogeneity in FFA, voxelwise tuning profiles were projected onto the first two PCs. Each point in these plots corresponds to a distinct voxel. Voxels assigned to distinct clusters by the spectral clustering analysis are annotated with a unique shape-color combination (see legend). Voxels in separate clusters have segregated distributions in the PC space as well. The first and second PCs distinguish voxels in the first and second clusters. In contrast, the second PC primarily distinguishes the third cluster from the rest. Therefore, the group structure revealed by the clustering analysis can be more confidently interpreted as a reflection of functional subdomains within FFA.

whether the dimensions of category information represented in FFA are similar to those represented broadly across occipitotemporal cortex. We reasoned that, if FFA is part of a broad representational map, the dimensions of category information represented in FFA should accurately explain category responses in occipitotemporal cortex. To examine this issue, we measured voxelwise category tuning profiles for a large set of voxels in occipitotemporal cortex. For each subject, this set included all voxels on the cortical surface within a 40 mm radius from the geometric center of FFA in each hemisphere. We then projected tuning profiles for the entire population of voxels extending well beyond FFA onto the first three group PCs obtained from FFA voxels alone.

We compared the PCs of FFA voxels with the PCs of the broader set of voxels in occipitotemporal cortex, in terms of the amount of response variance they explain across the occipitotemporal voxels. We find that the first three PCs of FFA voxels explain $42.2 \pm 1.5\%$ (mean $\pm$ SD across subjects) of the variance in category responses of voxels extending beyond FFA. In comparison, when PCA is directly performed on the entire population of occipitotemporal voxels within 40 mm from the center of FFA, the first three individual-subject PCs explain $49.2 \pm 1.5\%$ (mean $\pm$ SD across subjects) of the variance in category responses. This result indicates that FFA is embedded within a continuous map of category representation across a much broader region in occipitotemporal cortex.

To reveal the organization of category representation across the cortical surface, we visualized the PC projections on cortical flatmaps (Figs. 10 and Fig. 16). Voxels in areas along CoS (e.g., PPA) and the medial bank of inferior temporal sulcus (ITS) have positive projections onto the first PC. Voxels in areas along MTS (e.g., EBA) and the lateral bank of ITS have positive projections onto the second PC. Last, voxels in areas along ITS, located centrally within FFA, have positive projections onto the third PC. The projection patterns closely capture known functional properties of these areas in terms of category tuning (Huth et al., 2012). Therefore, these results support the view that FFA corre-

Figure 10. Cortical flatmaps of category tuning within and outside FFA. To examine the cortical distribution of category tuning within and outside FFA, voxelwise tuning profiles were projected onto the first three PCs measured from tuning profiles of FFA voxels (Fig. 8). Top row, Projections of tuning profiles onto the 3D space defined by PCs 1–3 are shown for subjects S1–S3. Projections onto the first, second, and third PCs determined the value of the red, green, and blue channels, respectively ($p < 0.05$, χ^2 test, FDR corrected). Here, voxels with similar category tuning have similar projections onto the PCs and are assigned similar colors (see legend). Insignificant voxels are shown in gray ($p > 0.05$). Category tuning is heterogeneously distributed within FFA. Individual FFA voxels are tuned for humans, artifacts, animals, gestures, and verbal communication. Bottom row, Projections of tuning profiles onto individual PCs are shown with separate cortical flatmaps for subject S2 (Fig. 15 for remaining subjects). Voxels with positive projections appear in red, and voxels with negative projections appear in blue (see legend). Brain areas identified using functional localizers are annotated with white lines and labeled (PPA; EBA; FBA, fusiform body area; OFA, occipital face area). Major sulci are denoted with black lines and labeled (CoS; ITS; MTS). Regions of fMRI signal dropout are indicated with dark gray patches. Voxels in areas along CoS (e.g., PPA) and the medial bank of ITS have positive projections onto the first PC. Voxels in areas along MTS (e.g., EBA) and the lateral bank of ITS have positive projections onto the second PC. Finally, voxels in areas along ITS, located centrally within FFA, have positive projections onto the third PC. These patterns suggest that broader gradients of category tuning across ventral–temporal cortex underlie the organization of the functional subdomains within FFA.

sponds to a focal peak along smooth gradients of category selectivity across occipitotemporal cortex.

Hemispheric lateralization

Many previous studies on FFA report that this area is predominantly larger in the right than in the left hemisphere (Sergent et al., 1992; Kanwisher and McDermott, 1997; Kanwisher and Yovel, 2006). It has even been suggested that right and left FFA might be specialized in different aspects of face processing (Rossion et al., 2000; Meng et al., 2012). We therefore asked whether the voxel clusters identified here exhibit any lateralization across hemispheres. To address this issue, we first counted the number of FFA voxels in the left and right hemispheres separately. We find that FFA is larger in the right hemisphere for two subjects, larger in the left hemisphere for one subject, and balanced across the hemispheres for the remaining two subjects. However, there is no consistent hemispheric lateralization in FFA definitions across subjects ($p > 0.30$, bootstrap test).

Next, we examined the distribution of voxels in each cluster between the left and right hemispheres. To account for intersubject differences in the size of FFA, the number of voxels in each cluster in each hemisphere was normalized by the total number of FFA voxels across both hemispheres. We find no evidence for hemispheric lateralization for any of the three clusters ($p > 0.34$, bootstrap test). This result indicates that the voxel clusters in FFA are symmetrically distributed across the hemispheres.

All clustering analyses presented up to this point were performed jointly on voxels within left and right FFA. A potential concern about this procedure is that it could underestimate tuning differences between cerebral hemispheres. To address this
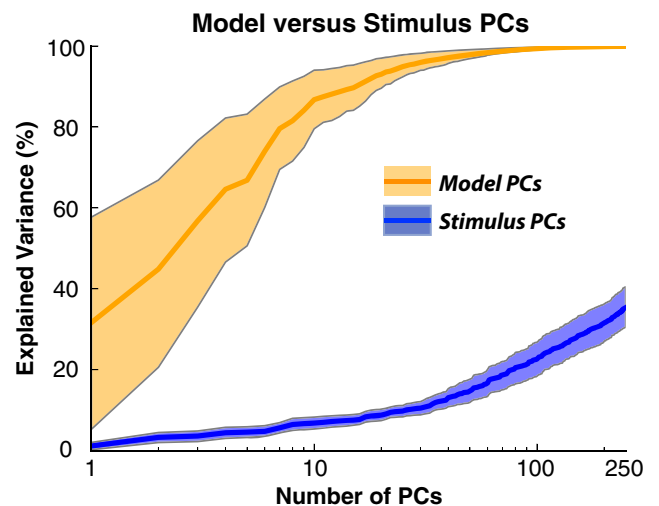
Figure 11. Temporal stimulus correlations between categories in natural movies. To assess whether category models are biased by stimulus correlations present in natural movies, PCs of category model weights were compared with stimulus PCs. To account for temporally lagged stimulus correlations, multiple delayed stimulus time courses were generated for each category using lags from -5 to 5 s. The delayed time courses for all categories were concatenated to form a stimulus matrix. Stimulus PCs were obtained from this matrix. Line plots show the mean explained variance and 68th percentile bands across the population of FFA voxels. Model PCs account for a significantly larger proportion of variance in the model weights compared with stimulus PCs ($p < 10^{-4}$, bootstrap test), regardless of the number of PCs used. This result indicates that the category model weights are not biased by category co-occurrences.

issue, we performed clustering analyses separately on FFA voxels in the left and right hemispheres of individual subjects. We find that the left and right hemisphere cluster centers are highly correlated with each other ($r = 0.73 \pm 0.16$, $p < 10^{-4}$, bootstrap test), and with the group cluster centers (0.84 ± 0.10, $p < 10^{-4}$). The within-cluster correlations are higher than the across-cluster correlations ($\Delta r = 0.17 \pm 0.04$, $p < 10^{-4}$). This result indicates that the voxel clusters are highly consistent across hemispheres.

Another concern related to voxel-selection procedures considers FFA definitions. Following standard procedures, here we defined FFA volumetrically by means of a localizer. The resulting FFA definitions can contain both gray and white matter voxels. To ensure that our results are not biased by differences between gray and white matter, we performed a separate clustering analysis after constraining FFA definitions to the cortical surface. We find that the cluster centers computed from surface-based and volume-based FFA definitions are highly correlated in all subjects ($r = 0.92 \pm 0.07$, $p < 10^{-4}$). Furthermore, the within-cluster correlations are higher than the across-cluster correlations ($\Delta r = 0.26 \pm 0.07$, $p < 10^{-4}$). This result suggests that functional heterogeneity in FFA is reliable regardless of voxel-selection procedures.
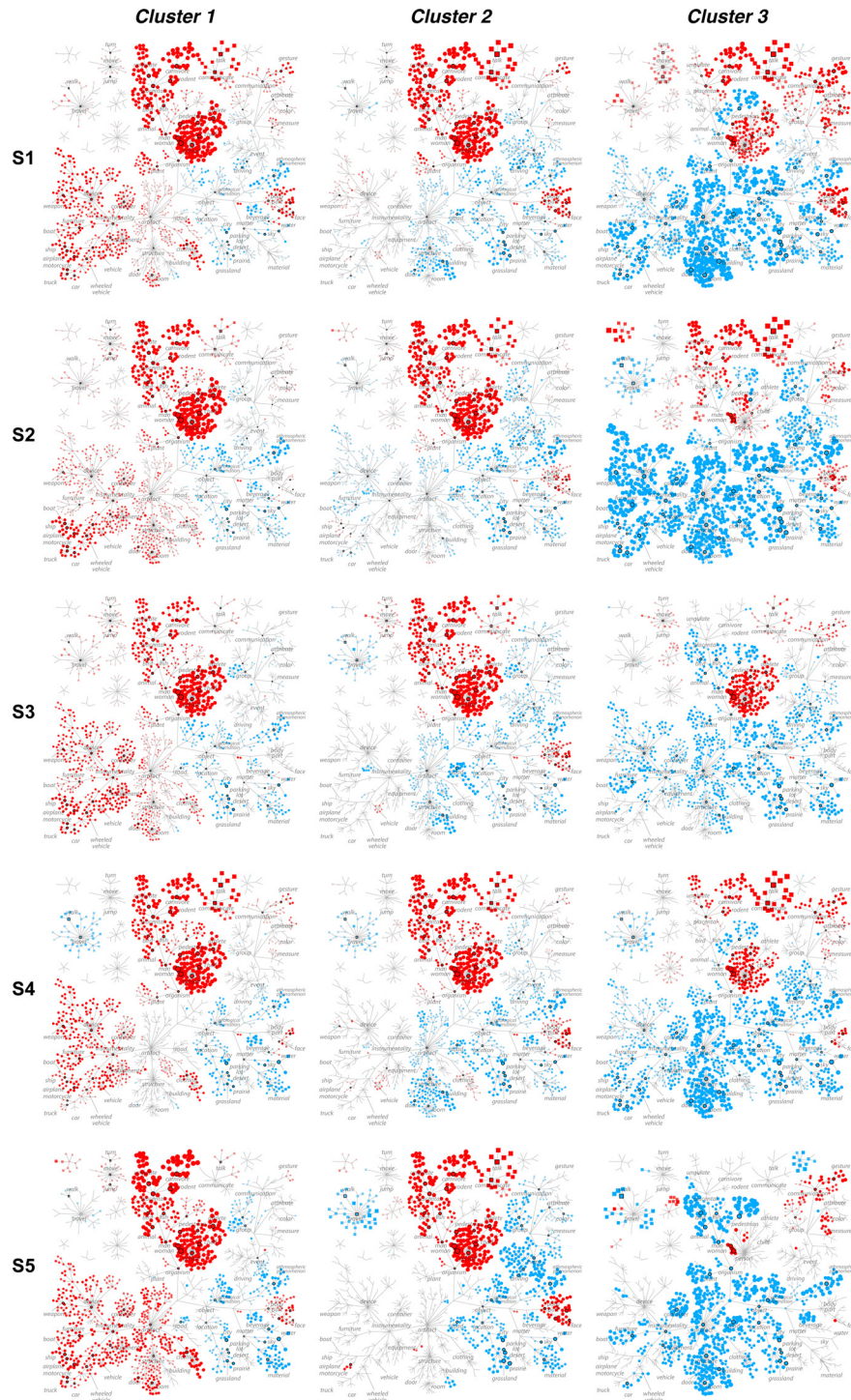
Stimulus correlations

One potential confound that could have influenced the results above concerns the potentially contaminating effects of stimulus correlations. Each frame of the natural movie stimulus used in our study contains multiple object or action categories. Stimulus correlations may arise when certain groups of categories tend to co-occur, either instantaneously or with a temporal lag. If such correlations exist, they might bias the category tuning profiles measured here toward broad tuning for multiple categories. To assess whether our results are biased by correlations present in our stimulus, we compared the PCs of category model weights with stimulus PCs. To account for temporally lagged stimulus correlations, multiple delayed stimulus time courses were generated for each category using lags from -5 to 5 s. The delayed time courses for all categories were concatenated to form a stimulus matrix, and stimulus PCs were obtained from this matrix. Figure 11 shows the proportion of variance in model weights explained by model PCs versus stimulus PCs. We find that model PCs account for a significantly larger proportion of variance for voxels in each individual cluster and subject ($p < 10^{-4}$, bootstrap test), regardless of the total number of PCs used. This result indicates that multicategory tuning in FFA is not an artifact of stimulus correlations at the level of categories.
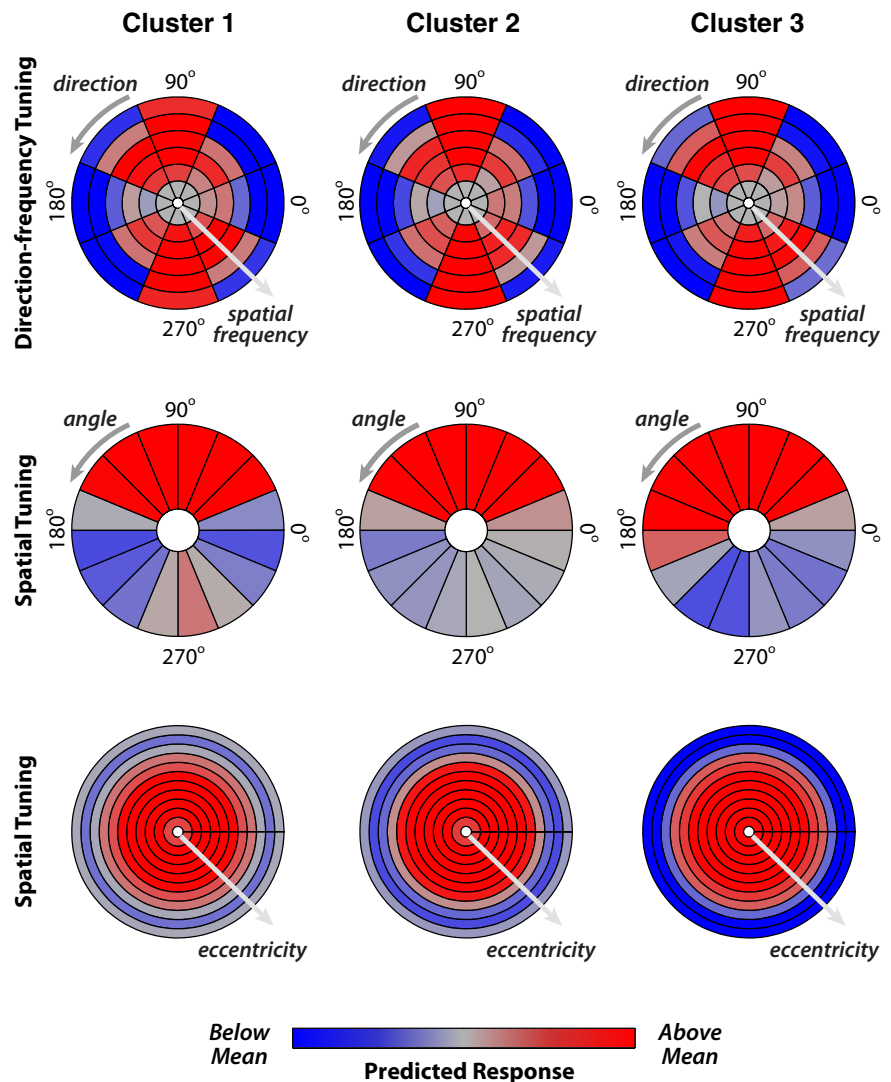
Figure 12. The cluster centers computed from the motion-energy model. The mean motion-energy tuning across voxels assigned to each cluster shown in Figure 5. The motion-energy tuning is shown with bull's-eye plots for a temporal frequency of 0 Hz. The tuning was nearly identical for other temporal frequencies (2, 4 Hz; data not shown). Top row, The radial component represents increasing spatial frequency (0, 2, 4, 8, 16, and 32 cycles/image), and the angular component represents direction (0 and 180 degrees indicate the horizontal direction). Middle row, The angular component represents polar angle. Bottom row, The radial component represents eccentricity (0, 1.2, 2.4 . . . 12.0 degrees). Red represents visual features that evoke above-mean responses; blue represents features that evoke below-mean responses. The mean motion-energy tuning is highly similar across the voxel clusters in FFA.

Another potential confound related to natural movies concerns stimulus correlations at the level of elementary visual features. Although FFA is thought to be primarily involved in category representation, several studies have suggested that elementary visual features are also represented in FFA (Fujita et al., 1992; Grill-Spector et al., 1999; Levy et al., 2001; Tsunoda et al., 2001). It is thus possible that multicategory tuning might emerge because of shared visual features between the stimulus categories that evoke responses in FFA. To ensure that our results are not confounded by such stimulus correlations, we fit separate voxel-wise motion-energy models to assess selectivity for elementary visual features. The motion-energy tuning profiles characterized selectivity for spatiotemporal frequency, direction, visual angle, and eccentricity. The mean motion-energy tuning profiles were computed within each cluster identified in the main analysis (i.e., based on category tuning profiles). Figure 12 displays the mean

Figure 13. Cluster centers of individual subjects. Mean tuning profiles across voxels within each cluster ($p < 0.05$, Monte Carlo test, FDR corrected) are shown separately for each individual subject (S1–S5). Formatting is identical to Figure 5.

strap test). This result indicates that multicategory tuning in FFA is not an artifact of stimulus correlations at the level of elementary visual features.
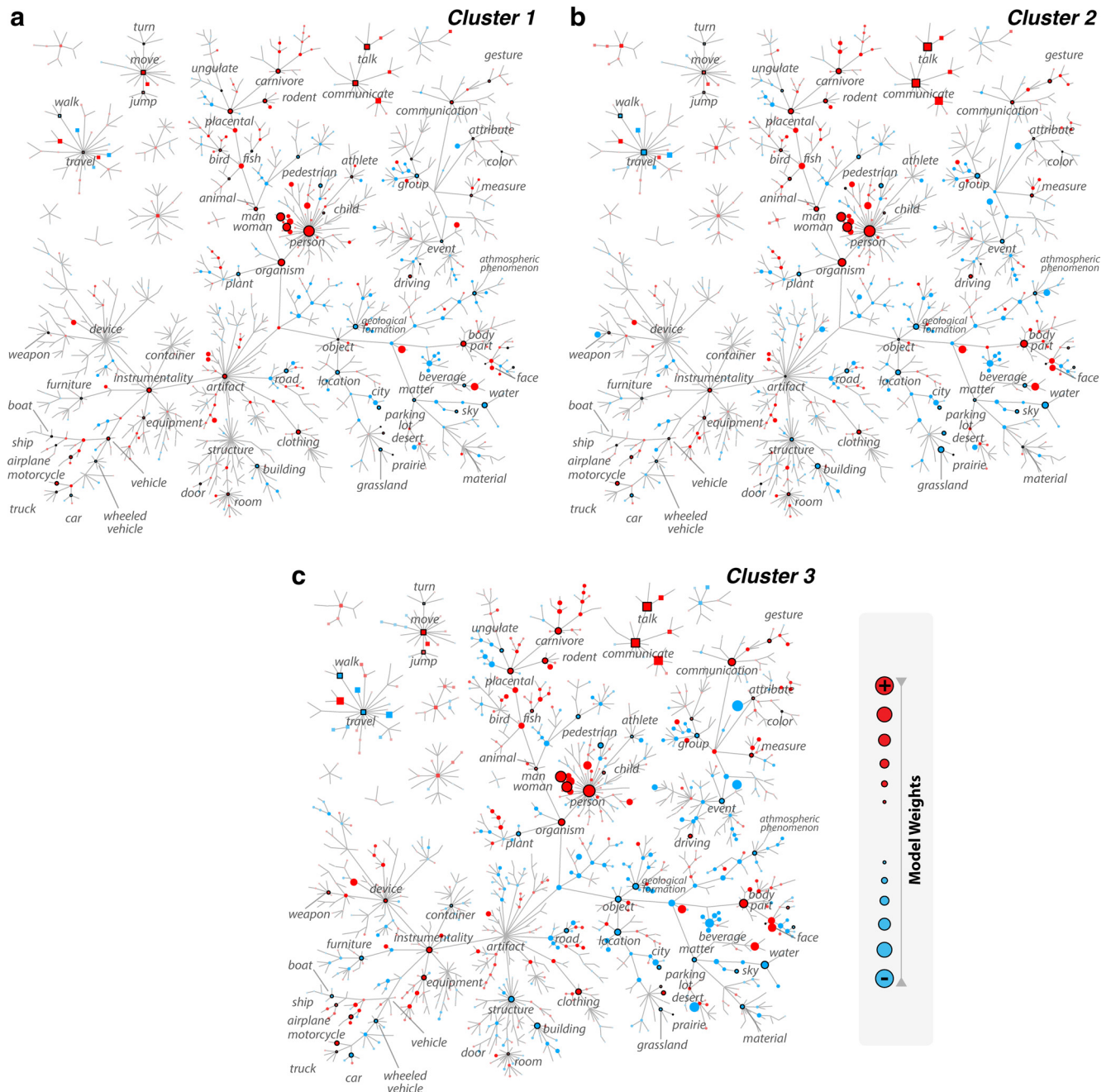
Discussion

We fit voxelwise encoding models in individual subjects using BOLD responses evoked by natural movies. These models characterize the tuning of FFA voxels for a large number of categories. Although all FFA voxels are strongly responsive to faces, these voxels are also differentially responsive to other categories, such as animals, vehicles, communication verbs, and body parts. A subsequent clustering analysis on the voxelwise tuning profiles reveals three spatially segregated functional subdomains within FFA of individual subjects. The main distinction among these subdomains is their tuning for the aforementioned nonface categories. We also find that heterogeneous category tuning is evident regardless of the p value threshold used to localize the FFA voxels.

Comparisons with previous studies

The results reported here are in disagreement with earlier studies that interpret FFA as an entirely homogeneous module (Puce et al., 1996; Kanwisher and McDermott, 1997; Spiridon and Kanwisher, 2002). Numerous aspects of experimental design and data analysis might have rendered these earlier studies less sensitive in detecting heterogeneous category tuning within FFA. In particular, a large volume of the earlier work was based on simple contrasts among a restricted number of categories (Weiner and Grill-Spector, 2012). This prohibits any inferences about categories that were not included in the stimulus set. Furthermore, significant tuning for nonface categories has often been disregarded based on a rather arbitrary threshold adapted from the neurophysiology literature (Perrett et al., 1982). As such, FFA has been designated as exclusively selective for a single category (i.e., faces) that elicits the maximum amount of BOLD responses. This predisposition may have led to the underestimation of tuning heterogeneity across FFA voxels.

Several recent fMRI studies have casted doubt on whether FFA can be viewed as an entirely homogeneous module. In particular, a group of studies used multivoxel pattern analyses to discriminate facial identities from responses measured in several regions near the fusiform gyrus, including FFA (Nestor et al., 2011; Goesaert and Op de Beeck, 2013). The results of these studies imply that individual faces elicit differential patterns of responses across FFA voxels. However, these previous studies did not examine the spatial organization face

motion-energy tuning for the clusters shown in Figure 5. The motion-energy tuning of all clusters is similar. To compare the heterogeneity of category versus motion-energy tuning, we quantified cluster separability as the average silhouette width across FFA voxels (see Materials and Methods). Positive widths indicate well-separated clusters, whereas negative widths indicate poorly separated clusters. We find that the average silhouette width of category tuning is significantly larger than that of motion-energy tuning in individual subjects ($\Delta S = 0.25 \pm 0.05$, $p < 10^{-4}$, boot-
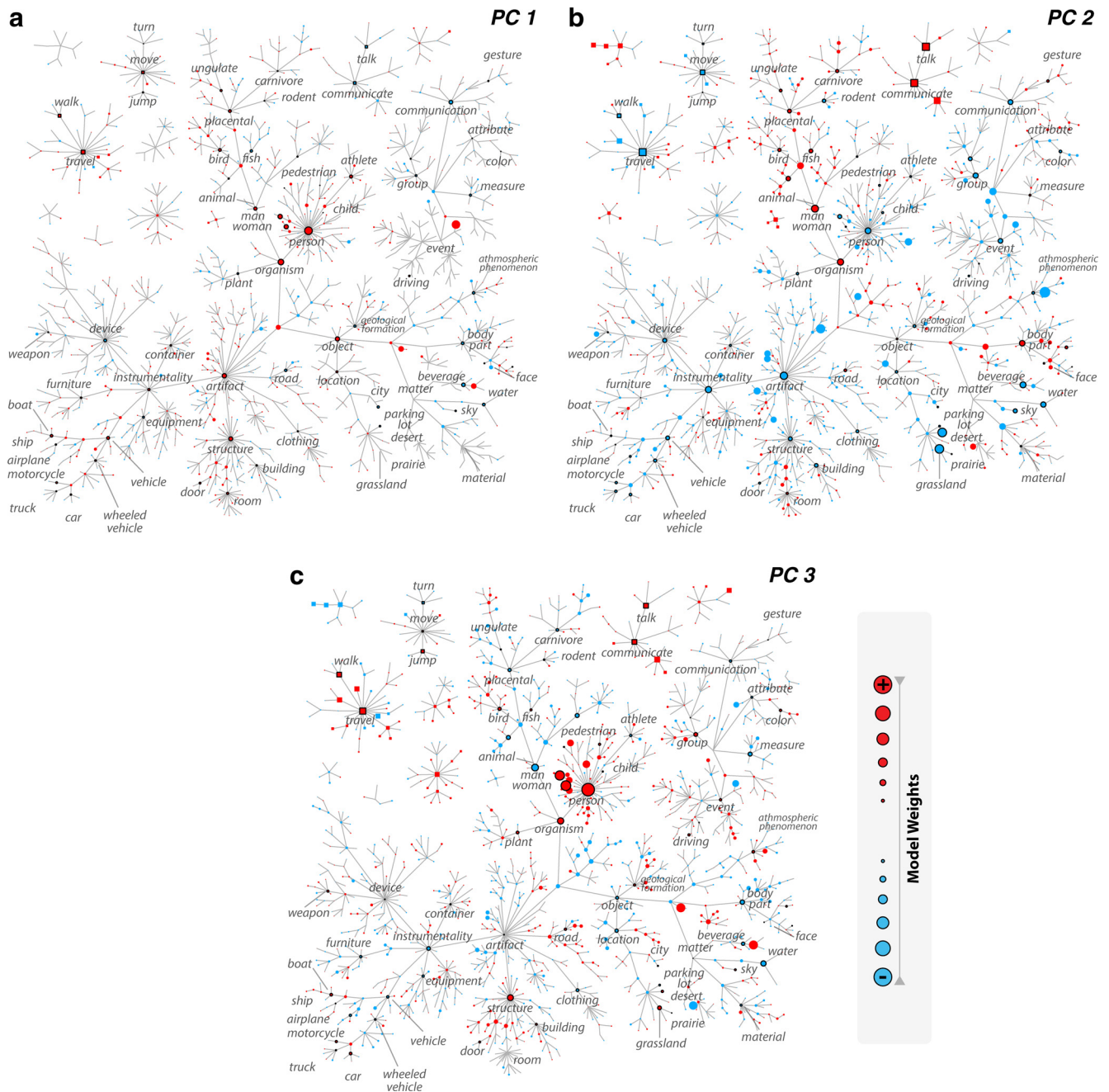
Figure 14. Group cluster centers. *a– c*, Mean raw model weights across voxels within each cluster (*p* < 0.05, Monte Carlo test, FDR corrected) are shown separately. Formatting is identical to Figure 5.

responses within FFA, and they did not focus on the representation of nonface categories. Thus, they did not make explicit predictions about the heterogeneity and the spatial organization of category tuning within FFA.

Although there have been previous reports of significant FFA responses for nonface categories (Gauthier et al., 1999; Ishai et al., 1999; Haxby et al., 2001), only few studies have attempted to assess the spatial distribution of category tuning within this region (Grill-Spector et al., 2006; Hanson and Schmidt, 2011). These high-spatial-resolution fMRI studies imply that regions within FFA might be tuned for nonface categories. In particular, the study by Hanson and Schmidt (2011) shows that the diagnostic weights for nonlinear classifiers trained to discriminate between several object categories (e.g., faces, animals, cars) exhibit a

heterogeneous distribution across FFA voxels. However, information required to discriminate between visual categories is implicitly present in retinal images. Therefore, a sufficiently strong nonlinear classifier could, in principle, perform category discrimination solely based on the activity of retinal ganglion cells. Analogously, nonlinear classifiers can recover category information from a brain area, even if none of the individual voxels within the area is tuned for visual categories. Thus, it is difficult to interpret the results obtained using nonlinear classifiers (Kamitani and Tong, 2005; Norman et al., 2006). In contrast, our results provide direct evidence for functional heterogeneity of FFA voxels apparent at typical resolutions prescribed in conventional fMRI studies.

Figure 15. Group PCs of FFA voxels. *a–c*, The raw model weights for the first three principal components of category tuning profiles ($p < 0.05$, Monte Carlo test, FDR corrected) are shown separately. Formatting is identical to Figure 8.

It has been argued that, if tuning for an object category is functionally important, then this tuning should be evident even when other categories are present in the visual environment (Kanwisher, 2010). Most previous studies on FFA that imply tuning for nonface categories used stimuli that consisted of single, isolated objects. In contrast, a recent study by Reddy and Kanwisher (2007) compared the level of information that FFA represents about nonface categories when the object of interest is displayed separately versus when it is displayed together with unwanted objects. Reddy and Kanwisher (2007) reported that the considerable level of information about nonface categories in isolated scenes reduces to insignificance in cluttered scenes. Here, we used a broad selection of natural movies that simultaneously depict multiple objects

under real-world viewing conditions. Yet, we measure broad tuning for nonface categories when using these highly complex stimuli. Therefore, our results provide strong evidence that FFA tuning for nonface categories is functionally important during natural vision.

Such discrepancies might reflect differences in the signal-to-noise ratio of various fMRI experiments. Previous studies of category representation have often used stimuli composed of flashed images rather than movies. In contrast, more reliable neural responses have been reported when presenting continuous movie clips (Yao et al., 2007) versus static stimuli (Tolhurst et al., 2009). Movie stimuli have also been shown to elicit large, synchronous BOLD responses in human fMRI experiments (Hasson et al., 2004). Furthermore, voxelwise encoding models have been
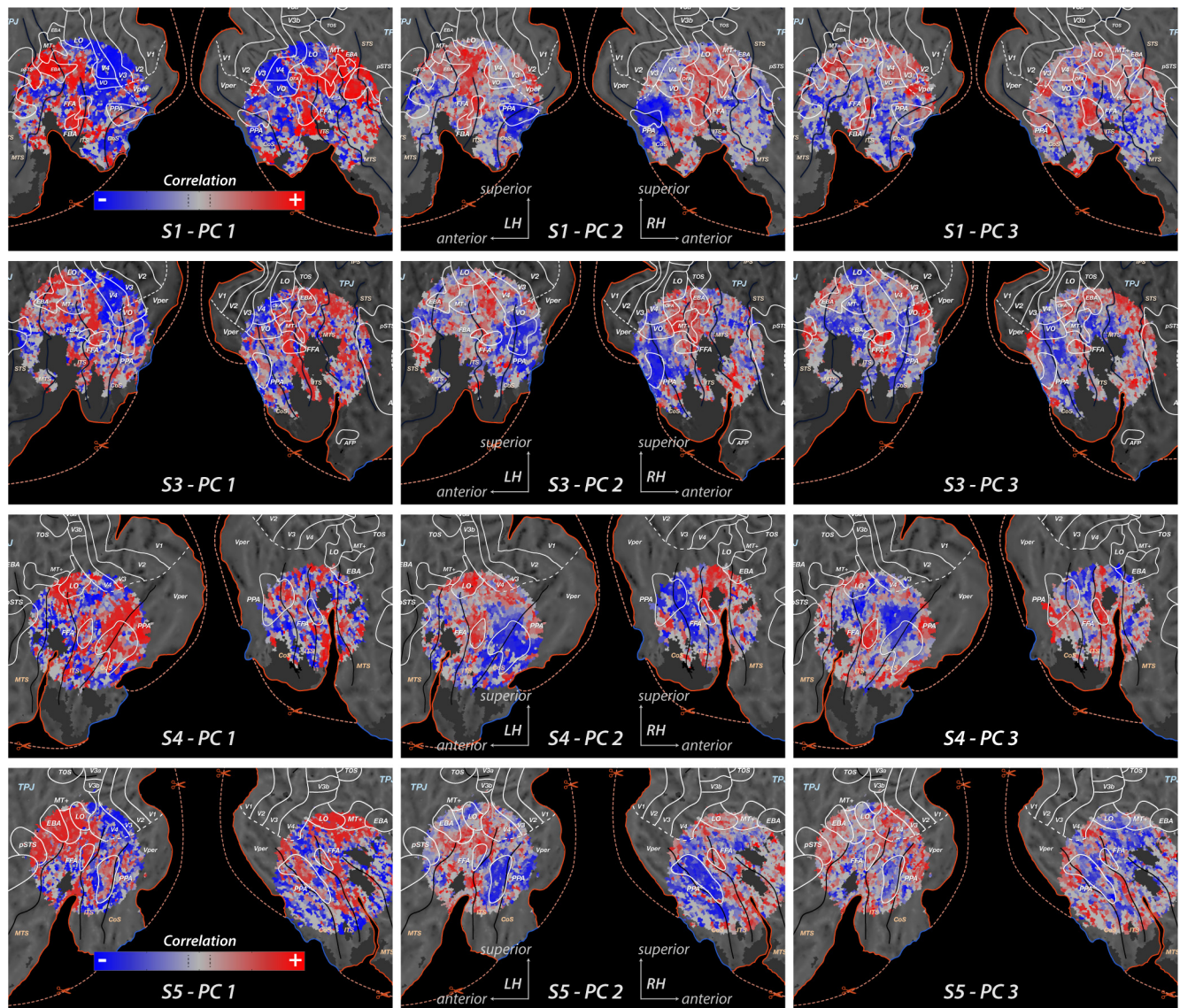
Figure 16. Cortical flatmaps of category tuning within and outside FFA. To examine the cortical distribution of category tuning within and outside FFA, voxelwise tuning profiles were projected onto the first three PCs measured from tuning profiles of FFA voxels (Fig. 8). Projections of tuning profiles onto individual PCs are shown on the cortical flatmaps for subjects S1, S3, S4, and S5 ($p <$ 0.05, χ^2 test, FDR corrected). Formatting is identical to Figure 10.

shown to predict responses to novel natural stimuli with unprecedented accuracy (Kay et al., 2008; Nishimoto et al., 2011). This modeling framework maximizes functional signal-to-noise ratio by using customized data preprocessing algorithms to compensate for various noise factors and by analyzing fMRI data for single voxels in the brain spaces of individual subjects (Sabuncu et al., 2010). Therefore, we think that our experimental procedures enable comprehensive testing of hypotheses on category representation with improved sensitivity compared with conventional methods.

Accounts for spatially heterogeneous category tuning
There are several plausible interpretations of heterogeneous tuning in FFA, all of which signify a more diverse role for FFA in shape coding than simple face detection. Previous studies have reported that growing expertise in identifying subordinate instances of objects within a category strengthens FFA responses (Gauthier et al., 1999, 2000; Golby et al., 2001). These data have been used to argue that category representation in FFA reflects

expertise in object recognition (Tarr and Gauthier, 2000). The clusters identified in our study exhibit differential tuning for nonface categories that commonly appear in daily life settings in which humans encounter faces. These typical settings include faces in civilized environments (involving vehicles, instruments, buildings etc.), faces of other humans and animals in isolation, and faces during social interaction with other humans (involving communication and body parts). Tuning for these nonface categories may provide contextual cues that facilitate recognition of faces under different settings (Bar, 2004; Rabinovich et al., 2007). Therefore, our results suggest that subregions of FFA may be specialized for detecting faces in ecologically relevant environments.

Alternatively, each of these clusters may reflect segregated regions in FFA that are involved in processing distinct object categories (Grill-Spector et al., 2006). For example, one cluster of voxels identified in our study exhibits strong tuning for both animals and humans. This result is in agreement with previous studies that reported a strong correlation between FFA responses

and behavioral performance in face and animal recognition tasks (Grill-Spector et al., 2004). Nonetheless, it is possible that there exist separate subpopulations of neurons within this voxel cluster that are selectively tuned for only animals or only faces (Grill-Spector et al., 2006). The aggregated activity of these neural populations may still exhibit multicategory tuning at the level of single voxels. Therefore, the level of functional heterogeneity reported here likely underestimates the true diversity of the underlying neural population.

The temporal cortex is commonly assumed to contain representations of visual objects (Haxby et al., 2001), but it is unclear which aspects of objects are represented where (Fujita et al., 1992; Grill-Spector et al., 1999; Levy et al., 2001; Tsunoda et al., 2001). Our results indicate that category tuning within FFA does not reflect simple selectivity for low-level features, such as spatial frequency or orientation. However, we cannot determine from our data whether FFA represents high-level object category information or mid-level structural features that are systematically related to faces (Op de Beeck et al., 2008). For instance, several previous studies have argued that FFA represents information about perceptual features related to facial identity (Gilaie-Dotan and Malach, 2007; Nestor et al., 2011; Harvey and Burgund, 2012; Goesaert and Op de Beeck, 2013). It is possible that the voxel clusters identified here are differentially tuned for such perceptual features and respond distinctively to specific face exemplars. Further research will be required to disentangle potential interactions between object categories and mid-level visual features.

In conclusion, our results provide direct evidence that FFA voxels are broadly tuned for both faces and several nonface categories, and that these voxels constitute a heterogeneous population in terms of their category tuning profiles. These findings support the hypothesis that object representation and information processing are not strictly constrained to specific modules but rather are distributed across the human cortex (Ishai et al., 1999; Haxby et al., 2001; Hanson et al., 2004). Therefore, our results have implications for brain regions in ventral–temporal cortex in addition to FFA that have traditionally been assumed to be category specific (Kanwisher, 2010). It will be an important topic for future research to reveal whether these other ROIs also exhibit tuning for multiple categories and whether they contain functional subdomains that differ in category tuning.

References

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617–629. CrossRef Medline

Ben-Hur A, Elisseeff A, Guyon I (2002) A stability based method for discovering structure in clustered data. In: Pacific Symposium on Biocomputing, pp 6–17. Lihue, HI: World Scientific: Pacific Symposium on Biocomputing.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188. CrossRef

DeAngelis GC, Ghose GM, Ohzawa I, Freeman RD (1999) Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. J Neurosci 19:4046–4064. Medline

Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N (2006) Domain specificity in visual cortex. Cereb Cortex 16:1453–1461. CrossRef Medline

Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex 7:181–192. CrossRef Medline

Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. Neuroimage 30:1077–1087. CrossRef Medline

Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. Nature 360:343–346. CrossRef Medline

Gallant JL, Nishimoto S, Naselaris T, Wu MC (2012) System identification,

encoding models and decoding models, a powerful new approach to fMRI. In: Visual population codes: towards a common multivariate framework for cell recording and functional imaging (Kriegeskorte N, Kreiman G, eds), pp 163–188. Boston: MIT.

Gauthier I (2000) What constrains the organization of the ventral temporal cortex? Trends Cogn Sci 4:1–2. CrossRef Medline

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci 2:568–573. CrossRef Medline

Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci 3:191–197. CrossRef Medline

Gilaie-Dotan S, Malach R (2007) Sub-exemplar shape tuning in human face-related areas. Cereb Cortex 17:325–338. CrossRef Medline

Goesaert E, Op de Beeck HP (2013) Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. J Neurosci 33:8549–8558. CrossRef Medline

Golby AJ, Gabrieli JD, Chiao JY, Eberhardt JL (2001) Differential responses in the fusiform region to same-race and other-race faces. Nat Neurosci 4:845–850. CrossRef Medline

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24:187–203. CrossRef Medline

Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 7:555–562. CrossRef Medline

Grill-Spector K, Sayres R, Ress D (2006) High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. Nat Neurosci 9:1177–1185. CrossRef Medline

Handl J, Knowles J, Kell DB (2005) Computational cluster validation in post-genomic data analysis. Bioinformatics 21:3201–3212. CrossRef Medline

Hansen KA, Kay KN, Gallant JL (2007) Topographic organization in and near human visual area V4. J Neurosci 27:11896–11911. CrossRef Medline

Hanson SJ, Schmidt A (2011) High-resolution imaging of the fusiform face area (FFA) using multivariate non-linear classifiers shows diagnosticity for non-face categories. Neuroimage 54:1715–1734. CrossRef Medline

Hanson SJ, Matsuka T, Haxby JV (2004) Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a 'face' area? Neuroimage 23:156–166. CrossRef Medline

Harvey DY, Burgund ED (2012) Neural adaptation across viewpoint and exemplar in fusiform cortex. Brain Cogn 80:33–44. CrossRef Medline

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. Science 303:1634–1640. CrossRef Medline

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430. CrossRef Medline

Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ (2011) A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72:404–416. CrossRef Medline

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210–1224. CrossRef Medline

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV (1999) Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci U S A 96:9379–9384. CrossRef Medline

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270.

Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841. CrossRef Medline

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685. CrossRef Medline

Kanwisher N (2010) Functional specificity in the human brain: a window into the functional architecture of the mind. Proc Natl Acad Sci U S A 107:11163–11170. CrossRef Medline

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a

module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311. Medline

Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci 361:2109 –2128. CrossRef Medline

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352–355. CrossRef Medline

Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center-periphery organization of human object areas. Nat Neurosci 4:533–539. Medline

Luxburg U (2007) A tutorial on spectral clustering. Stat Comput 17:395– 416. CrossRef

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability, pp 281–297. Berkeley, CA: University of California.

Meng M, Cherian T, Singal G, Sinha P (2012) Lateralization of face processing in the human brain. Proc Biol Sci 279:2052–2061. CrossRef Medline

Miller G (1995) WordNet: a lexical database for English. Commun ACM 38:39–41. CrossRef

Mur M, Ruff DA, Bodurka J, De Weerd P, Bandettini PA, Kriegeskorte N (2012) Categorical, yet graded, single-image activation profiles of human category-selective cortical regions. J Neurosci 32:8649 –8662. CrossRef Medline

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. Neuron 63: 902–915. CrossRef Medline

Nestor A, Plaut DC, Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. Proc Natl Acad Sci U S A 108:9998 –10003. CrossRef Medline

Ng AY, Jordan MI, Weiss Y (2001) On spectral clustering: analysis and an algorithm. In: Advances in neural information processing systems (Dietterich T, Becker S, Ghahramani Z, eds), pp 849–856. Cambridge, MA: MIT.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21:1641–1646. CrossRef Medline

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci (Regul Ed) 10:424–430. CrossRef Medline

Op de Beeck HP, Haushofer J, Kanwisher NG (2008) Interpreting fMRI data: maps, modules and dimensions. Nat Rev Neurosci 9:123–135. CrossRef Medline

O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. J Cogn Neurosci 17:580–590. CrossRef Medline

Perrett DI, Rolls ET, Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329–342. Medline

Peter JR (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. CrossRef

Puce A, Allison T, Asgari M, Gore JC, McCarthy G (1996) Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J Neurosci 16:5205–5215. Medline

Rabinovich A, Vedaldi A, Galleguillos C, Wiewiora E, Belongie S (2007) Objects in context. In: Computer vision, 2007. ICCV 2007. IEEE 11th International Conference, Rio De Janeiro, pp 1–8. IEEE.

Reddy L, Kanwisher N (2007) Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr Biol 17: 2067–2072. CrossRef Medline

Rossion B, Dricot L, Devolder A, Bodart JM, Crommelinck M, De Gelder B, Zoontjes R (2000) Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. J Cogn Neurosci 12:793–802. CrossRef Medline

Sabuncu MR, Singer BD, Conroy B, Bryan RE, Ramadge PJ, Haxby JV (2010) Function-based intersubject alignment of human cortical anatomy. Cereb Cortex 20:130–140. CrossRef Medline

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893. CrossRef Medline

Sergent J, Ohta S, MacDonald B (1992) Functional neuroanatomy of face and object processing: a positron emission tomography study. Brain 115: 15–36. CrossRef Medline

Spiridon M, Kanwisher N (2002) How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron 35: 1157–1165. CrossRef Medline

Spiridon M, Fischl B, Kanwisher N (2006) Location and spatial profile of category-specific regions in human extrastriate cortex. Hum Brain Mapp 27:77–89. CrossRef Medline

Tarr MJ, Gauthier I (2000) FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat Neurosci 3:764–769. CrossRef Medline

Tolhurst DJ, Smyth D, Thompson ID (2009) The sparseness of neuronal responses in ferret primary visual cortex. J Neurosci 29:2355–2370. CrossRef Medline

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M (2001) Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat Neurosci 4:832–838. CrossRef Medline

Weiner KS, Grill-Spector K (2012) The improbable simplicity of the fusiform face area. Trends Cogn Sci (Regul Ed) 16:251–254. CrossRef Medline

Yao H, Shi L, Han F, Gao H, Dan Y (2007) Rapid learning in cortical coding of visual scenes. Nat Neurosci 10:772–778. CrossRef Medline

Zelnik-Manor L, Perona P (2004) Self-tuning spectral clustering. Adv Neural Information Processing Syst 17:1601–1608.